(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification[7]: C07K 14/315

(21) International Application Number:
PCT/EP2004/003984

(22) International Filing Date: 15 April 2004 (15.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
03450087.6 15 April 2003 (15.04.2003) EP

(71) Applicant (for all designated States except US): INTER-CELL AG [AT/AT]; Campus Vienna Biocenter 6, A-1030 Vienna (AT).

(72) Inventors; and
(75) Inventors/Applicants (for US only): MEINKE, Andreas [DE/AT]; Piettegasse 26/1, A-3013 Pressbaum (AT). NAGY, Eszter [HU/AT]; Taborstrasse 9, A-1020 Vienna (AT). HANNER, Markus [AT/AT]; Jacquingasse 5/6, A-1030 Vienna (AT). DEWASTHALY, Shailesh [IN/AT]; Anschützgasse 23/10, A-1150 Vienna (AT). STIER-SCHNEIDER, Ulrike [AT/AT]; Kaiserstrasse 66/2/14, A-1070 Vienna (AT).

(74) Agent: SONN & PARTNER; Riemergasse 14, A-1010 Vienna (AT).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations
— of inventorship (Rule 4.17(iv)) for US only

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: S. PNEUMONIAE ANTIGENS

(57) Abstract: The present invention discloses isolated nucleic acid molecules encoding a hyperimmune serum reactive antigen or a fragment thereof as well as hyperimmune serum reactive antigens or fragments thereof from S. pneumoniae, methods for isolating such antigens and specific uses thereof.

S.PNEUMONIAE ANTIGENS

The present invention relates to isolated nucleic acid molecules, which encode antigens for *Streptococcus pneumoniae*, which are suitable for use in preparation of pharmaceutical medicaments for the prevention and treatment of bacterial infections caused by *Streptococcus pneumoniae*.

Streptococcus pneumoniae (Pneumococcus) is a lancet-shaped, gram-positive, facultative anaerobic bacterium. It is only the encapsulated organism that is pathogenic for humans and experimental animals. Capsules are antigenic and form the basis for classifying pneumococci by serotypes. Ninety serotypes have been identified, based on their reaction with type-specific antisera. Most *S. pneumoniae* serotypes have been shown to cause serious disease, and the ten most common serotypes are estimated to account for about 62% of invasive disease worldwide. The ranking and serotype prevalence differs by age group and geographic area.

Pneumococci are common inhabitants of the respiratory tract, and may be isolated from the nasopharynx of 5% to 70% of normal adults. Rates of asymptomatic carriage vary with age, environment, and the presence of upper respiratory infections. Only 5%-10% of adults without children are carriers. In schools and orphanages, 27% to 58% of students and residents may be carriers. On military installations, as many as 50% to 60% of service personnel may be carriers. The duration of carriage varies and is generally longer in children than adults (reviewed in Epidemiology and Prevention of Vaccine-Preventable Diseases, 7th Edition-Second Printing, The Pink Book).

The relationship of carriage to the development of natural immunity is poorly understood. In addition, the immunologic mechanism that allows disease to occur in a carrier is poorly understood.

Streptococcus pneumoniae is an important agent of human disease at the extremities of age and in those who have underlying disease. Pneumococcal disease kills more people – in the US 40,000 or more each year – than all other vaccine preventable diseases combined. The major clinical syndromes of pneumococcal disease include pneumonia, bacteremia, and meningitis. The disease most often occurs when a predisposing condition exists, particularly pulmonary disease. It is a common bacterial complication of antecedent viral respiratory infection such as influenza and measles, and of chronic conditions such as chronic obstructive pulmonary disease, diabetes, congestive heart failure, renal failure, smoking and alcoholism. Pneumococcal infections are more common during the winter and in early spring when respiratory diseases are more prevalent. Immunodeficiency (splenic dysfunction, iatrogen, etc.) is a risk factor for development of fatal pneumococcal infections, because of decreased bacterial clearance and lack of antibodies. The incubation period is short, 1-3 days. Symptomes include an abrupt onset of fever and shaking chills or rigor, productive cough, pleuritic chest pain, dyspnoe, tachycardia and hypoxia.

S. pneumoniae is responsible for 88% of bacteremia infections in the US. Pneumonia is the most common form of invasive pneumococcal diseases: 150.000-570.000 cases per year (US). 36% of adult community-acquired and 50% of hospital-acquired pneumonia is caused by *S. pneumoniae* (US). The incidence of disease among adults aged 65 years and older has been reported to be ~60 cases/100.000. Case fatality rates for this disease increase from 1.4% for those aged two or younger to as high as 20.6% among those aged 80 or older. Diseases caused by influenza and *Pneumococcus* are together the fifth leading cause of death for persons aged 65 and older. Mortality attributable to these pathogens is more than 90% in this age group. Bacteremia occurs in about 25-30% of patients with pneumonia. The overall mortality rate of bacteremia is about 20%, but may be as high as 60% in elderly people. In 1998, 51% of all deaths attributable to invasive pneumococcal diseases occurred in age group above 65 years. Pneumococci cause 13%-19% of all cases of bacterial meningitis in the United States. An estimated 3,000 to 6,000 cases of pneumococcal meningitis occur each year. One-quarter of patients with pneumococcal meningitis also have pneumonia. The clinical symptoms, spinal fluid profile and neurologic complications are similar to other forms of purulent bacterial meningitis (reviewed in Epidemiology and Prevention of Vaccine-Preventable Diseases, 7th Edition-Second Printing, The Pink Book).

CONFIRMATION COPY

- 2 -

In children, Pneumococci are a common cause of acute otitis media, and are detected in 28%-55% of middle ear aspirates. By age 12 months, 62% of children have had at least one episode of acute otitis media. Middle ear infections are the most frequent reasons for pediatric office visits in the United States, resulting in over 20 million visits annually. Complications of pneumococcal otitis media may include mastoiditis and meningitis. Bacteremia without a known site of infection is the most common invasive clinical presentation among children <2 years of age, accounting for approximately 70% of invasive disease in this age group. Bacteremic pneumonia accounts for 12%-16% of invasive pneumococcal disease among children <2 years of age. With the decline of invasive Hib disease, *S. pneumoniae* has become the leading cause of bacterial meningitis among children <5 years of age in the United States. Children <1 year have the highest rates of pneumococcal meningitis, approximately 10 cases per 100,000 population. The burden of pneumococcal disease among children <5 years of age is significant. An estimated 17,000 cases of invasive disease occur each year, of which 13,000 are bacteremia without a known site of infection and about 700 are of meningitis. An estimated 200 children die every year as a result of invasive pneumococcal disease. Although not considered invasive disease, an estimated 5 million cases of acute otitis media occur each year among children <5 years of age (reviewed Epidemiology and Prevention of Vaccine-Preventable Diseases, 7th Edition-Second Printing, The Pink Book).

A definitive diagnosis of infection with *Streptococcus pneumoniae* generally relies on isolation of the organism from blood or other normally sterile body sites. Tests are also available to detect capsular polysaccharide antigen in body fluids.

Penicillin is the drug of choice for treatment. However, successful implementation of anti-infective therapy has become increasingly difficult because of widespread antimicrobial resistance. Resistance to penicillin is rising, and according to recent reports it reaches ~ 25% in the US {Whitney, C. et al., 2000}. The proportion of macrolide-resistant strains reached ~ 20 % {Hyde, T. et al., 2001}. Use of antimicrobial agents is highly correlated with the increase in resistance of *S. pneumoniae* to ɵ-lactams and macrolides {McCormick, A. et al., 2003}.

However, even with effective antibiotic therapy (sensitive strains), the case fatality rate of invasive disease is high with an average of 10% in the developed world and can be much higher with certain serotypes, in elderly patients and in cases of bacteremia or meningitis (up to 80%).

Thus, there remains a need for an effective treatment to prevent or ameliorate spneumoococcal infections. A vaccine could not only prevent infections by streptococci, but more specifically prevent or ameliorate colonization of host tissues (esp. in nasopharynx), thereby reducing the incidence of upper respiratory infections and other suppurative infections, such otitis media. Elimination of invasive diseases - pneumonia, bacteremia and meningitis, and sepsis - would be a direct consequence of reducing the incidence of acute infection and carriage of the organism. Vaccines capable of showing cross-protection against the majority of *S. pneumoniae* strains causing human infections would also be useful to prevent or ameliorate infections caused by all other streptococcal species, namely groups A, B, C and G.

A vaccine can contain a whole variety of different antigens. Examples of antigens are whole-killed or attenuated organisms, subfractions of these organisms/tissues, proteins, or, in their most simple form, peptides. Antigens can also be recognized by the immune system in form of glycosylated proteins or peptides and may also be or contain polysaccharides or lipids. Short peptides can be used since for example cytotoxic T-cells (CTL) recognize antigens in form of short usually 8-11 amino acids long peptides in conjunction with major histocompatibility complex (MHC). B-cells can recognize linear epitopes as short as 4-5 amino acids, as well as three-dimensional structures (conformational epitopes). In order to obtain sustained, antigen-specific immune responses, adjuvants need to trigger immune cascades that involve all cells of the immune system. Primarily, adjuvants are acting, but are not restricted in their mode of action, on so-called antigen presenting cells (APCs). These cells usually first

encounter the antigen(s) followed by presentation of processed or unmodified antigen to immune effector cells. Intermediate cell types may also be involved. Only effector cells with the appropriate specificity are activated in a productive immune response. The adjuvant may also locally retain antigens and co-injected other factors. In addition the adjuvant may act as a chemoattractant for other immune cells or may act locally and/or systemically as a stimulating agent for the immune system.

Efforts to develop effective pneumococcal vaccines began as early as 1911. However, with the advent of penicillin in the 1940s, interest in the vaccine declined, until it was observed that many patients still died despite antibiotic treatment. By the late 60s, efforts were again being made to develop a polyvalent vaccine. The first pneumococcal vaccines contained purified capsular polysaccharide antigen from 14 different types of pneumococcal bacteria. In 1983, a 23-valent polysaccharide vaccine (PPV23) was licensed and replaced the 14- valent vaccine, which is no longer produced. PPV23 contains polysaccharide antigen from 23 types of pneumococcal bacteria which cause 88% of bacteremic pneumococcal disease. In addition, cross-reactivity occurs for several capsular types which account for an additional 8% of bacteremic disease. Two polysaccharide vaccines are available in the United States (Pneumovax 23, Merck, and Pnu-Immune 23, Wyeth-Lederle). Both vaccines contain 25 μg of each antigen per dose and include either phenol or thimerosal as a preservative.

The first pneumococcal conjugate vaccine (PCV7, Prevnar) was licensed in the United States in 2000. It includes purified capsular polysaccharide of 7 serotypes of S. pneumoniae (4, 9V, 14, 19F, 23F, 18C, and 6B) conjugated to a nontoxic variant of diphtheria toxin known as CRM197. The serotypes included in Prevnar accounted for 86% of bacteremia, 83% of meningitis, and 65% of acute otitis media among children <6 years of age in the United States during 1978-1994 (reviewed in Epidemiology and Prevention of Vaccine-Preventable Diseases, 7th Edition-Second Printing, The Pink Book). Additional pneumococcal polysaccharide conjugate vaccines containing 9 and 11 serotypes of S. pneumoniae are being developed. The vaccine is administered intramuscularly. After 4 doses of Prevnar vaccine, virtually all healthy infants develop antibody to all 7 serotypes contained in the vaccine. Prevnar has also been shown to be immunogenic in infants and children, including those with sickle cell disease and HIV infection. In a large clinical trial, Prevnar was shown to reduce invasive disease caused by vaccine serotypes, and reduce invasive disease caused by all serotypes, including serotypes not in the vaccine. Children who received Prevnar had fewer episodes of acute otitis media and underwent fewer tympanostomy tube placements than unvaccinated children. The duration of protection following Prevnar is currently unknown. Immunization with Prevnar reduces the rate of nasopharyngeal carriage of the vaccine serotypes, while the overall carriage rate is unaffected. Unfortunately, it has also been shown to induce serotype redistribution, that is the replacement of vaccine serotypes by strains, which are not covered by Prevnar {Pelton, S. et al., 2003}.

Pneumococcal vaccine is recommended to be administered routinely to i., all children as part of the routine shildhood immunization schedule, ii., adults 65 years of age and older and iii., persons aged >2 years with normal immune systems who have chronic illnesses, including cardiovascular disease, pulmonary disease, diabetes, alcoholism, cirrhosis, or cerebrospinal fluid leaks. In the elderly population the target groups for pneumococcal vaccine and influenza vaccine overlap. These vaccines can be given at the same time at different sites without increased side effects.

High mortality is observed among high-risk individuals (with underlying disease – mainly viral respiratory infection, immunocompromise) even with effective antibiotic therapy. The mAb approach targets patients with serious disease and provides immediate immune enhancement for the clearance of the bacteria. Through opsonization bacteria are killed within phagocytic cells and not lysed in the blood by antibiotics. This mechanism of action can help to eliminate the release of toxins (such as pneumolysin and other cytotoxins), which worsen the clinical condition of septic patients. Recent advances in the technology of monoclonal antibody production provide the means to generate human antibody reagents and reintroduce antibody therapies, while avoiding the toxicities associated with serum therapy.

- 4 -

Immunoglobulins are an extremely versatile class of antimicrobial proteins that can be used to prevent and treat emerging infectious diseases. Antibody therapy has been effective against a variety of diverse microorganisms reviewed in {Burnie, J. et al., 1998}.

Although capsular specific antibodies have been shown to be highly protective, it remains unclear what concentration of these serotype-specific antibodies protect against disease and more recently it has become clear that opsonic activity and avidity of these antibodies are more critical determinants of protection than concentration.

Protein conjugate vaccines are no doubt a great new addition to the amarmatorium in the battle against pneumococcal disease, but the vaccine contains a limited number of pneumococcal serotypes and given adequate ecological pressure, replacement disease by non-vaccine serotypes remains a real threat, particularly in areas with very high disease burden.

During the last decade the immunogenicity and protective capacity of several pneumococcal proteins have been described in animal models and these are now being explored for the development of species-common protein based vaccines. Such proteins are the Pneumococcal surface protein A (PspA, {McDaniel, L. et al., 1991}; {Roche, H. et al., 2003}), Pneumococcal surface adhesin A (PsaA, {Talkington, D. et al., 1996}), Choline binding protein A (CbpA, {Rosenow, C. et al., 1997}), LytB glucosaminidase, LytC muramidase, PrtA serine protease, PhtA (histidine triad A) and Pneumococcal vaccine antigen A (PvaA) {Wizemann, T. et al., 2001}; {Adamou, J. et al., 2001}.

Certain proteins or enzymes displayed on the surface of gram-positive organisms significantly contribute to pathogenesis, and might be involved in the disease process caused by these pathogens. Often, these proteins are involved in direct interactions with host tissues or in conceiling the bacterial surface from the host defense mechanisms {Navarre, W. et al., 1999}. S. pneumoniae is not an exception in this regard. Several surface proteins are characterized by as virulence factors, important for pneumococcal pathogenicity reviewed in {Jedrzejas, M., 2001}. If antibodies to these proteins could offer better protection to humans, they could provide the source of a novel, protein-based pneumococcal vaccine to be used in conjunction with or in place of the more traditional capsular polysaccharise vaccine. The use of some of the above-described proteins as antigens for a potential vaccine as well as a number of additional candidates reviewed in {Di Guilmi, A. et al., 2002} resulted mainly from a selection based on easiness of identification or chance of availability. There is a demand to identify relevant antigens for S. pneumoniae in a more comprehensive way.

The present inventors have developed a method for identification, isolation and production of hyperimmune serum reactive antigens from a specific pathogen, especially from Staphylococcus aureus and Staphylococcus epidermidis (WO 02/059148). However, given the differences in biological property, pathogenic function and genetic background, Streptococcus pneumoniae is distinctive from Staphylococcus strains. Importantly, the selection of sera for the identification of antigens from S. pneumoniae is different from that applied to the S. aureus screens. Three major types of human sera were collected for that purpose. First, healthy adults below <45 years of age preferably with small children in the household were tested for nasopharyngeal carriage of S. pneumoniae. A large percentage of young children are carriers of S. pneumoniae, and they are considered to be a source for exposure for their family members. Based on correlative data, protective (colonization neutralizing) antibodies are likely to be present in exposed individuals (children with high carriage rate in the household) who are not carriers of S. pneumoniae. To be able to select for relevant serum sources, a series of ELISAs measuring anti-S. pneumoniae IgG and IgA antibody levels were performed with bacterial lysates and culture supernatant proteins. Sera from high titer non-carriers were included in the genomic-based antigen identification. This approach for selection of human sera is basically very different from that used for S. aureus, where carriage or non-carriage state couldn't be associated with antibody levels. Second, serum samples from

- 5 -

convalescent phase patients with invasive pneumococcal diseases were characterized and selected in the same way. The third group of sera, containing longitudinally collected samples were also obtained from individuals with invasive disease and were used mainly for validation purposes. The main value of this collection is that one can follow the changes in antigen-specific antibody levels before diase (prae-), at the time of onset (acute) and during recovery (convalescent). This latter group helps in the selection of epitopes, which induce antibodies during disease and missing in the prae-disease state.

The genomes of the two bacterial species *S. pneumoniae* and *S. aureus* by itself show a number of important differences. The genome of *S. pneumoniae* contains app. 2.16 Mb, while *S. aureus* harbours 2.85 Mb. They have an average GC content of 39.7 and 33%, respectively and approximately 30 to 45% of the encoded genes are not shared between the two pathogens. In addition, the two bacterial species require different growth conditions and media for propagation. While *S. pneumoniae* is a strictly human pathogen, *S. aureus* can also be found infecting a range of warm-blooded animals. A list of the most important diseases, which can be inflicted by the two pathogens is presented below. *S. aureus* causes mainly nosocomial, opportunistic infections: impetigo, folliculitis, abscesses, boils, infected lacerations, endocarditis, meningitis, septic arthritis, pneumonia, osteomyelitis, scalded skin syndrome (SSS), toxic shock syndrome. *S. pneumoniae* causes mainly community aquired infections: upper (pharyngitis, otitis media) and and lower respiratory infections (pneumonia), as well as bacteremia, sepsis and meningitis.

The complete genome sequence of a capsular serotype 4 isolate of *S. pneumoniae*, designated TIGR4 was determined by the random shotgun sequencing strategy (GenBank accession number AE005672; see www.tigr.org/tigrscripts/CMR2/CMRHomePage.spl). This clinical isolate was taken from the blood of a 30-year-old male patient in Kongsvinger, Norway, and is highly invasive and virulent in a mouse model of infection.

The problem underlying the present invention was to provide means for the development of medicaments such as vaccines against *S. pneumoniae* infection. More particularly, the problem was to provide an efficient, relevant and comprehensive set of nucleic acid molecules or hyperimmune serum reactive antigens from *S. pneumoniae* that can be used for the manufacture of said medicaments.

Therefore, the present invention provides an isolated nucleic acid molecule encoding a hyperimmune serum reactive antigen or a fragment thereof comprising a nucleic acid sequence, which is selected from the group consisting of:
   a) a nucleic acid molecule having at least 70% sequence identity to a nucleic acid molecule selected from Seq ID No 1, 101-144.
   b) a nucleic acid molecule which is complementary to the nucleic acid molecule of a),
   c) a nucleic acid molecule comprising at least 15 sequential bases of the nucleic acid molecule of a) or b)
   d) a nucleic acid molecule which anneals under stringent hybridisation conditions to the nucleic acid molecule of a), b), or c)
   e) a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the nucleic acid molecule defined in a), b), c) or d).

According to a preferred embodiment of the present invention the sequence identity is at least 80%, preferably at least 95%, especially 100%.

Furthermore, the present invention provides an isolated nucleic acid molecule encoding a hyperimmune serum reactive antigen or a fragment thereof comprising a nucleic acid sequence selected from the group consisting of
   a) a nucleic acid molecule having at least 96% sequence identity to a nucleic acid molecule selected from Seq ID No 2-6, 8, 10-16, 18-23, 25-31, 34, 36, 38-42, 44, 47-48, 51, 53, 55-62, 64, 67, 71-76, 78-79, 81-94, 96-100.

- 6 -

b) a nucleic acid molecule which is complementary to the nucleic acid molecule of a),

c) a nucleic acid molecule comprising at least 15 sequential bases of the nucleic acid molecule of a) or b)

d) a nucleic acid molecule which anneals under stringent hybridisation conditions to the nucleic acid molecule of a), b) or c),

e) a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the nucleic acid defined in a), b), c) or d).

According to another aspect, the present invention provides an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of

a) a nucleic acid molecule selected from Seq ID No 9, 17, 24, 32, 37, 43, 52, 54, 65-66, 70, 80.

b) a nucleic acid molecule which is complementary to the nucleic acid of a),

c) a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the nucleic acid defined in a), b), c) or d).

Preferably, the nucleic acid molecule is DNA or RNA.

According to a preferred embodiment of the present invention, the nucleic acid molecule is isolated from a genomic DNA, especially from a S. *pneumoniae* genomic DNA.

According to the present invention a vector comprising a nucleic acid molecule according to any of the present invention is provided.

In a preferred embodiment the vector is adapted for recombinant expression of the hyperimmune serum reactive antigens or fragments thereof encoded by the nucleic acid molecule according to the present invention.

The present invention also provides a host cell comprising the vector according to the present invention.

According to another aspect the present invention further provides a hyperimmune serum-reactive antigen comprising an amino acid sequence being encoded by a nucleic acid molecule according to the present invention.

In a preferred embodiment the amino acid sequence (polypeptide) is selected from the group consisting of Seq ID No 145, 245-288.

In another preferred embodiment the amino acid sequence (polypeptide) is selected from the group consisting of Seq ID No 146-150, 152, 154-160, 162-167, 169-175, 178, 180, 182-186, 188, 191-192, 195, 197, 199-206, 208, 211, 215-220, 222-223, 225-238, 240-244.

In a further preferred embodiment the amino acid sequence (polypeptide) is selected from the group consisting of Seq ID No 153, 161, 168, 176, 181, 187, 196, 198, 209-210, 214, 224.

According to a further aspect the present invention provides fragments of hyperimmune serum-reactive antigens selected from the group consisting of peptides comprising amino acid sequences of column "predicted immunogenic aa" and "location of identified immunogenic region" of Table 1; the serum reactive epitopes of Table 2, especially peptides comprising amino acids 4-11, 35-64, 66-76, 101-108, 111-119 and 57-114 of Seq ID No 145; 5-27, 32-64, 92-102, 107-113, 119-125, 133-139, 148-162, 177-187, 195-201, 207-214, 241-251, 254-269, 285-300, 302-309, 317-324, 332-357, 365-404, 411-425, 443-463, 470-477, 479-487, 506-512, 515-520, 532-547, 556-596, 603-610, 616-622, 624-629, 636-642, 646-665, 667-674, 687-692, 708-720, 734-739, 752-757, 798-820, 824-851, 856-865 and 732-763 of Seq ID No 146; 14-21, 36-44, 49-66, 102-127, 162-167, 177-196, 45-109 and 145-172 of Seq ID No 147; 17-35, 64-75, 81-92, 100-119, 125-172, 174-183, 214-

222, 230-236, 273-282, 287-303, 310-315, 331-340, 392-398, 412-420, 480-505, 515-523, 525-546, 553-575, 592-598, 603-609, 617-625, 631-639, 644-651, 658-670, 681-687, 691-704, 709-716, 731-736, 739-744, 750-763, 774-780, 784-791, 799-805, 809-822, 859-870, 880-885, 907-916, 924-941, 943-949, 973-986, 1010-1016, 1026-1036, 1045-1054, 1057-1062, 1082-1088, 1095-1102, 1109-1120, 1127-1134, 1140-1146, 1152-1159, 1169-1179, 1187-1196, 1243-1251, 1262-1273, 1279-1292, 1306-1312, 1332-1343, 1348-1364, 1379-1390, 1412-1420, 1427-1436, 1458-1468, 1483-1503, 1524-1549, 1574-1588, 1614-1619, 1672-1685, 1697-1707, 1711-1720, 1738-1753, 1781-1787, 1796-1801, 1826-1843, 132-478, 508-592 and 1753-1810 of Seq ID No 148; 15-43, 49-55, 71-77, 104-110, 123-130, 162-171, 180-192, 199-205, 219-227, 246-254, 264-270, 279-287, 293-308, 312-322, 330-342, 349-356, 369-377, 384-394, 401-406, 416-422, 432-439, 450-460, 464-474, 482-494, 501-508, 521-529, 536-546, 553-558, 568-574, 584-591, 602-612, 616-626, 634-646, 653-660, 673-681, 688-698, 705-710, 720-726, 736-749, 833-848, 1-199, 200-337, 418-494 and 549-647 of Seq ID No 149; 9-30, 65-96, 99-123, 170-178 and 1-128 of Seq ID No 150; 7-32, 34-41, 96-106, 127-136, 154-163, 188-199, 207-238, 272-279, 306-312, 318-325, 341-347, 353-360, 387-393, 399-406, 434-440, 452-503, 575-580, 589-601, 615-620, 635-640, 654-660, 674-680, 696-701, 710-731, 1-548 and 660-691 of Seq ID No 151; 4-19, 35-44, 48-59, 77-87, 93-99, 106-111, 130-138, 146-161 and 78-84 of Seq ID No 152; 24-30, 36-43, 64-86, 93-99, 106-130, 132-145, 148-165, 171-177, 189-220, 230-249, 251-263, 293-300, 302-312, 323-329, 338-356, 369-379, 390-412 and 179-193 of Seq ID No 153; 30-39, 61-67, 74-81, 90-120, 123-145, 154-167, 169-179, 182-197, 200-206, 238-244, 267-272 and 230-265 of Seq ID No 154; 14-20, 49-65, 77-86 and 2-68 of Seq ID No 155; 4-9, 26-35, 42-48, 53-61, 63-85, 90-101, 105-111, 113-121, 129-137, 140-150, 179-188, 199-226, 228-237, 248-255, 259-285, 299-308, 314-331, 337-343, 353-364, 410-421, 436-442 and 110-144 of Seq ID No 156; 36-47, 55-63, 94-108, 129-134, 144-158, 173-187, 196-206, 209-238, 251-266, 270-285, 290-295, 300-306, 333-344, 346-354, 366-397, 404-410, 422-435, 439-453, 466-473, 515-523, 529-543, 554-569, 571-585, 590-596, 607-618, 627-643, 690-696, 704-714, 720-728, 741-749, 752-767, 780-799, 225-247 and 480-507 of Seq ID No 157; 16-25, 36-70, 80-93, 100-106 and 78-130 of Seq ID No 158; 18-27, 41-46, 50-57, 65-71, 79-85, 93-98, 113-128, 144-155, 166-178, 181-188, 201-207, 242-262, 265-273, 281-295, 303-309, 318-327 and 36-64 of Seq ID No 159; 7-29, 31-44, 50-59, 91-96, 146-153, 194-201, 207-212, 232-238, 264-278, 284-290, 296-302, 326-353, 360-370, 378-384, 400-405, 409-418, 420-435, 442-460, 499-506, 529-534, 556-562, 564-576, 644-651, 677-684, 687-698, 736-743, 759-766, 778-784, 808-814, 852-858, 874-896, 920-925, 929-935, 957-965, 1003-1012, 1021-1027, 1030-1044, 1081-1087, 1101-1111, 1116-1124, 1148-1159, 1188-1196, 1235-1251, 1288-1303, 1313-1319, 1328-1335, 1367-1373, 1431-1437, 1451-1458, 1479-1503, 1514-1521, 1530-1540, 1545-1552, 1561-1568, 1598-1605, 1617-1647, 1658-1665, 1670-1676, 1679-1689, 1698-1704, 1707-1713, 1732-1738, 1744-1764, 1-70, 154-189, 922-941, 1445-1462 and 1483-1496 of Seq ID No 160; 6-51, 81-91, 104-113, 126-137, 150-159, 164-174, 197-209, 215-224, 229-235, 256-269, 276-282, 307-313, 317-348, 351-357, 376-397, 418-437, 454-464, 485-490, 498-509, 547-555, 574-586, 602-619 and 452-530 of Seq ID No 161; 25-31, 39-47, 49-56, 99-114, 121-127, 159-186, 228-240, 253-269, 271-279, 303-315, 365-382, 395-405, 414-425, 438-453 and 289-384 of Seq ID No 162; 9-24, 41-47, 49-54, 68-78, 108-114, 117-122, 132-140, 164-169, 179-186, 193-199, 206-213, 244-251, 267-274, 289-294, 309-314, 327-333, 209-249 and 286-336 of Seq ID No 163; 9-28, 53-67, 69-82, 87-93, 109-117, 172-177, 201-207, 220-227, 242-247, 262-268, 305-318, 320-325 and 286-306 of Seq ID No 164; 4-10, 26-39, 47-58, 63-73, 86-96, 98-108, 115-123, 137-143, 148-155, 160-176, 184-189, 194-204, 235-240, 254-259, 272-278 and 199-283 of Seq ID No 165; 4-26, 33-39, 47-53, 59-65, 76-83, 91-97, 104-112, 118-137, 155-160, 167-174, 198-207, 242-268, 273-279, 292-315, 320-332, 345-354, 358-367, 377-394, 403-410, 424-439, 445-451, 453-497, 511-518, 535-570, 573-589, 592-601, 604-610 and 202-242 of Seq ID No 166; 8-30, 36-45, 64-71, 76-82, 97-103, 105-112, 134-151, 161-183, 211-234, 253-268, 270-276, 278-284, 297-305, 309-315, 357-362, 366-372, 375-384, 401-407, 409-416, 441-455, 463-470, 475-480, 490-497, 501-513, 524-537, 552-559, 565-576, 581-590, 592-600, 619-625, 636-644, 646-656 and 316-419 of Seq ID No 167; 4-17, 52-58, 84-99, 102-110, 114-120, 124-135, 143-158, 160-173, 177-196, 201-216, 223-250, 259-267, 269-275 and 1-67 of Seq ID No 168; 6-46, 57-67, 69-80, 82-133, 137-143, 147-168, 182-187, 203-209, 214-229, 233-242, 246-280 and 53-93 of Seq ID No 169; 7-40, 50-56, 81-89, 117-123, 202-209, 213-218, 223-229, 248-261, 264-276, 281-288, 303-308, 313-324, 326-332, 340-346, 353-372, 434-443, 465-474, 514-523, 556-564, 605-616, 620-626, 631-636, 667-683, 685-699, 710-719, 726-732, 751-756, 760-771, 779-788, 815-828, 855-867, 869-879, 897-902, 917-924, 926-931, 936-942, 981-1000, 1006-1015, 1017-1028, 1030-1039, 1046-1054, 1060-1066, 1083-1092, 1099-1112, 1122-1130, 1132-1140, 1148-1158, 1161-1171, 1174-1181, 1209-1230, 1236-1244, 1248-1254, 1256-1267, 1269-1276, 1294-1299, 1316-1328, 1332-1354, 1359-1372, 1374-1380, 1384-1390, 1395-1408, 1419-1425, 1434-1446, 1453-1460, 1465-1471, 1474-

1493, 1505-1515, 1523-1537, 1547-1555, 1560-1567, 1577-1605, 1633-1651, 1226-1309, 1455-1536 and 1538-1605 of Seq ID No 170; 4-10, 31-39, 81-88, 106-112, 122-135, 152-158, 177-184, 191-197, 221-227, 230-246, 249-255, 303-311, 317-326, 337-344, 346-362, 365-371, 430-437, 439-446, 453-462, 474-484 and 449-467 of Seq ID No 171; 9-15, 24-35, 47-55, 122-128, 160-177, 188-196, 202-208, 216-228, 250-261, 272-303, 318-324, 327-339, 346-352, 355-361, 368-373, 108-218 and 344-376 of Seq ID No 172; 6-14, 17-48, 55-63, 71-90, 99-109, 116-124, 181-189, 212-223, 232-268, 270-294, 297-304, 319-325, 340-348, 351-370, 372-378, 388-394, 406-415, 421-434 and 177-277 of Seq ID No 173; 21-39, 42-61, 65-75, 79-85, 108-115 and 11-38 of Seq ID No 174; 4-17, 26-39, 61-76, 103-113, 115-122, 136-142, 158-192, 197-203, 208-214, 225-230, 237-251 and 207-225 of Seq ID No 175; 5-11, 27-36, 42-53, 62-70, 74-93, 95-104, 114-119, 127-150, 153-159, 173-179, 184-193, 199-206, 222-241, 248-253, 257-280, 289-295, 313-319, 322-342, 349-365, 368-389, 393-406, 408-413, 426-438, 447-461, 463-470, 476-495, 532-537, 543-550 and 225-246 of Seq ID No 176; 4-29, 68-82, 123-130, 141-147, 149-157, 178-191, 203-215, 269-277, 300-307, 327-335, 359-370, 374-380, 382-388, 393-400, 410-417, 434-442, 483-492, 497-503, 505-513, 533-540, 564-569, 601-607, 639-647, 655-666, 693-706, 712-718, 726-736, 752-758, 763-771, 774-780, 786-799, 806-812, 820-828, 852-863, 884-892, 901-909, 925-932, 943-948, 990-996, 1030-1036, 1051-1059, 1062-1068, 1079-1086, 1105-1113, 1152-1162, 1168-1179, 1183-1191, 1204-1210, 1234-1244, 1286-1295, 1318-1326, 1396-1401, 1451-1460, 1465-1474, 1477-1483, 1488-1494, 1505-1510, 1514-1521, 1552-1565, 1593-1614, 1664-1672, 1677-1685, 1701-1711, 1734-1745, 1758-1770, 1784-1798, 1840-1847, 1852-1873, 1885-1891, 1906-1911, 1931-1939, 1957-1970, 1977-1992, 2014-2020, 2026-2032, 2116-2134, 1-348, 373-490, 573-767, 903-1043, 1155-1198, 1243-1482, 1550-1595, 1682-1719, 1793-1921 and 2008-2110 of Seq ID No 177; 10-35, 39-52, 107-112, 181-188, 226-236, 238-253, 258-268, 275-284, 296-310, 326-338, 345-368, 380-389, 391-408, 410-418, 420-429, 444-456, 489-505, 573-588, 616-623, 637-643, 726-739, 741-767, 785-791, 793-803, 830-847, 867-881, 886-922, 949-956, 961-980, 988-1004, 1009-1018, 1027-1042, 1051-1069, 1076-1089, 1108-1115, 1123-1135, 1140-1151, 1164-1179, 1182-1191, 1210-1221, 1223-1234, 1242-1250, 1255-1267, 1281-1292, 1301-1307, 1315-1340, 1348-1355, 1366-1373, 1381-1413, 1417-1428, 1437-1444, 1453-1463, 1478-1484, 1490-1496, 1498-1503, 1520-1536, 1538-1546, 1548-1570, 1593-1603, 1612-1625, 1635-1649, 1654-1660, 1670-1687, 1693-1700, 1705-1711, 1718-1726, 1729-1763, 1790-1813, 1871-1881, 1893-1900, 1907-1935, 1962-1970, 1992-2000, 2006-2013, 2033-2039, 2045-2051, 2055-2067, 2070-2095, 2097-2110, 2115-2121, 2150-2171, 2174-2180, 2197-2202, 2206-2228 and 1526-1560 of Seq ID No 178; 4-17, 35-48, 54-76, 78-107, 109-115, 118-127, 134-140, 145-156, 169-174, 217-226, 232-240, 256-262, 267-273, 316-328, 340-346, 353-360, 402-409, 416-439, 448-456, 506-531, 540-546, 570-578, 586-593, 595-600, 623-632, 662-667, 674-681, 689-705, 713-724, 730-740, 757-763, 773-778, 783-796, 829-835, 861-871, 888-899, 907-939, 941-955, 957-969, 986-1000, 1022-1028, 1036-1044, 1068-1084, 1095-1102, 1118-1124, 1140-1146, 1148-1154, 1168-1181, 1185-1190, 1197-1207, 1218-1226, 1250-1270, 1272-1281, 1284-1296, 1312-1319, 1351-1358, 1383-1409, 1422-1428, 1438-1447, 1449-1461, 1482-1489, 1504-1510, 1518-1527, 1529-1537, 1544-1551, 1569-1575, 1622-1628, 1631-1637, 1682-1689, 1711-1718, 1733-1740, 1772-1783, 1818-1834, 1859-1872, 1-64 and 128-495 of Seq ID No 179; 8-28, 32-37, 62-69, 119-125, 137-149, 159-164, 173-189, 200-205, 221-229, 240-245, 258-265, 268-276, 287-293, 296-302, 323-329 and 1-95 of Seq ID No 180; 9-18, 25-38, 49-63, 65-72, 74-81, 94-117, 131-137, 139-146, 149-158, 162-188, 191-207, 217-225, 237-252, 255-269, 281-293, 301-326, 332-342, 347-354, 363-370, 373-380, 391-400, 415-424, 441-447 and 75-107 of Seq ID No 181; 4-24, 64-71, 81-87, 96-116, 121-128, 130-139, 148-155, 166-173, 176-184, 203-215, 231-238, 243-248, 256-261, 280-286, 288-306, 314-329 and 67-148 of Seq ID No 182; 4-10, 19-37, 46-52, 62-81, 83-89, 115-120, 134-139, 141-151, 168-186, 197-205, 209-234, 241-252, 322-335, 339-345, 363-379, 385-393, 403-431, 434-442, 447-454, 459-465, 479-484, 487-496 and 404-420 of Seq ID No 183; 10-35, 46-66, 71-77, 84-93, 96-122, 138-148, 154-172, 182-213, 221-233, 245-263, 269-275, 295-301, 303-309, 311-320, 324-336, 340-348, 351-359, 375-381 and 111-198 of Seq ID No 184; 14-25, 30-42, 47-61, 67-75, 81-91, 98-106, 114-122, 124-135, 148-193, 209-227 and 198-213 of Seq ID No 185; 5-18, 45-50, 82-90, 97-114, 116-136, 153-161, 163-171, 212-219, 221-227, 240-249, 267-281, 311-317, 328-337, 375-381, 390-395, 430-436, 449-455, 484-495, 538-543, 548-554, 556-564, 580-586, 596-602 and 493-606 of Seq ID No 186; 9-25, 28-34, 37-44, 61-68, 75-81, 88-96, 98-111, 119-133, 138-150, 152-163, 168-182, 186-194, 200-205, 216-223, 236-245, 257-264, 279-287, 293-304, 311-318, 325-330, 340-346, 353-358, 365-379, 399-409, 444-453 and 303-391 of Seq ID No 187; 16-36, 55-61, 66-76, 78-102, 121-130, 134-146, 150-212, 221-239, 255-276, 289-322, 329-357 and 29-59 of Seq ID No 188; 8-27, 68-74, 77-99, 110-116, 124-141, 171-177, 202-217, 221-228, 259-265, 275-290, 293-303, 309-325, 335-343, 345-351, 365-379, 384-394, 406-414, 423-437, 452-465, 478-507, 525-534, 554-560, 611-624, 628-651, 669-682, 742-747, 767-778, 782-

792, 804-812, 820-836, 79-231 and 359-451 of Seq ID No 189; 5-28, 39-45, 56-62, 67-74, 77-99, 110-117, 124-141, 168-176, 200-230, 237-244, 268-279, 287-299, 304-326, 329-335, 348-362, 370-376, 379-384, 390-406, 420-429, 466-471, 479-489, 495-504, 529-541, 545-553, 561-577, 598-604, 622-630, 637-658, 672-680, 682-688, 690-696, 698-709, 712-719, 724-736, 738-746, 759-769, 780-786, 796-804, 813-818, 860-877, 895-904, 981-997, 1000-1014, 1021-1029, 1-162, 206-224, 254-350, 414-514 and 864-938 of Seq ID No 190; 4-11, 19-49, 56-66, 68-101, 109-116, 123-145, 156-165, 177-185, 204-221, 226-234, 242-248, 251-256, 259-265, 282-302, 307-330, 340-349, 355-374, 377-383, 392-400, 422-428, 434-442, 462-474 and 266-322 of Seq ID No 191; 14-43, 45-57, 64-74, 80-87, 106-127, 131-142, 145-161, 173-180, 182-188, 203-210, 213-219, 221-243, 245-254, 304-311, 314-320, 342-348, 354-365, 372-378, 394-399, 407-431, 436-448, 459-465, 470-477, 484-490, 504-509, 531-537, 590-596, 611-617, 642-647, 723-734, 740-751, 754-762, 764-774, 782-797, 807-812, 824-831, 838-845, 877-885, 892-898, 900-906, 924-935, 940-946, 982-996, 1006-1016, 1033-1043, 1051-1056, 1058-1066, 1094-1108, 1119-1126, 1129-1140, 1150-1157, 1167-1174, 1176-1185, 1188-1201, 1209-1216, 1220-1228, 1231-1237, 1243-1248, 1253-1285, 1288-1297, 1299-1307, 1316-1334, 1336-1343, 1350-1359, 1365-1381, 1390-1396, 1412-1420, 1427-1439, 1452-1459, 1477-1484, 1493-1512, 1554-1559, 1570-1578, 1603-1608, 1623-1630, 1654-1659, 1672-1680, 1689-1696, 1705-1711, 1721-1738, 1752-1757, 1773-1780, 1817-1829, 1844-1851, 1856-1863, 1883-1895, 1950-1958, 1974-1990, 172-354, 384-448, 464-644, 648-728 and 1357-1370 of Seq ID No 192; 8-27, 68-74, 77-99, 110-116, 124-141, 169-176, 201-216, 220-227, 258-264, 274-289, 292-302, 308-324, 334-342, 344-350, 364-372, 377-387, 399-407, 416-429, 445-458, 471-481, 483-500, 518-527, 547-553, 604-617, 621-644, 662-675, 767-778, 809-816, 15-307, 350-448 and 496-620 of Seq ID No 193; 4-17, 24-29, 53-59, 62-84, 109-126, 159-164, 189-204, 208-219, 244-249, 274-290, 292-302, 308-324, 334-342, 344-350, 378-389, 391-397, 401-409, 424-432, 447-460, 470-479, 490-504, 521-529, 538-544, 549-555, 570-577, 583-592, 602-608, 615-630, 635-647, 664-677, 692-698, 722-731, 733-751, 782-790, 793-799, 56-267, 337-426 and 495-601 of Seq ID No 194; 12-22, 49-59, 77-89, 111-121, 136-148, 177-186, 207-213, 217-225, 227-253, 259-274, 296-302, 328-333, 343-354, 374-383, 424-446, 448-457, 468-480, 488-502, 507-522, 544-550, 553-560, 564-572, 587-596, 604-614, 619-625, 629-635, 638-656, 662-676, 680-692, 697-713, 720-738, 779-786, 833-847, 861-869, 880-895, 897-902, 911-917, 946-951, 959-967, 984-990, 992-1004, 1021-1040, 1057-1067, 1073-1080 and 381-403 of Seq ID No 195; 4-10, 26-31, 46-56, 60-66, 70-79, 86-94, 96-102, 109-118, 132-152, 164-187, 193-206, 217-224 and 81-149 of Seq ID No 196; 4-21, 26-37, 48-60, 71-82, 109-117, 120-128, 130-136, 142-147, 181-187, 203-211, 216-223, 247-255, 257-284, 316-325, 373-379, 395-400, 423-435, 448-456, 479-489, 512-576, 596-625, 641-678, 680-688, 692-715 and 346-453 of Seq ID No 197; 10-16, 25-31, 34-56, 58-69, 71-89, 94-110, 133-176, 186-193, 208-225, 240-250, 259-266, 302-307, 335-341, 376-383, 410-416 and 316-407 of Seq ID No 198; 11-29, 42-56, 60-75, 82-88, 95-110, 116-126, 132-143, 145-160, 166-172, 184-216 and 123-164 of Seq ID No 199; 11-29, 54-63, 110-117, 139-152, 158-166, 172-180, 186-193, 215-236, 240-251, 302-323, 330-335, 340-347, 350-366, 374-381 and 252-299 of Seq ID No 200; 18-27, 35-42, 50-56, 67-74, 112-136, 141-153, 163-171, 176-189, 205-213, 225-234, 241-247, 253-258, 269-281, 288-298, 306-324, 326-334, 355-369, 380-387 and 289-320 of Seq ID No 201; 7-15, 19-41, 56-72, 91-112, 114-122, 139-147, 163-183, 196-209, 258-280, 326-338, 357-363, 391-403, 406-416 and 360-378 of Seq ID No 202; 11-18, 29-41, 43-49, 95-108, 142-194, 204-212, 216-242, 247-256, 264-273 and 136-149 of Seq ID No 203; 18-24, 33-40, 65-79, 89-102, 113-119, 130-137, 155-161, 173-179, 183-203, 205-219, 223-231, 245-261, 267-274, 296-306, 311-321, 330-341, 344-363, 369-381, 401-408, 415-427, 437-444, 453-464, 472-478, 484-508, 517-524, 526-532, 543-548 and 59-180 of Seq ID No 204; 5-13, 52-65, 67-73, 97-110, 112-119, 134-155 and 45-177 of Seq ID No 205; 6-28, 34-43, 57-67, 75-81, 111-128, 132-147, 155-163, 165-176, 184-194, 208-216, 218-229, 239-252, 271-278, 328-334, 363-376, 381-388, 426-473, 481-488, 492-498, 507-513, 536-546, 564-582, 590-601, 607-623, 148-269, 420-450 and 610-648 of Seq ID No 206; 4-12, 20-38, 69-75, 83-88, 123-128, 145-152, 154-161, 183-188, 200-213, 245-250, 266-272, 306-312, 332-339, 357-369, 383-389, 395-402, 437-453, 455-470, 497-503 and 1-112 of Seq ID No 207; 35-59, 74-86, 111-117, 122-137 and 70-154 of Seq ID No 208; 26-42, 54-61, 65-75, 101-107, 123-130, 137-144, 148-156, 164-172, 177-192, 213-221, 231-258 and 157-249 of Seq ID No 209; 29-38, 61-67, 77-87, 94-100, 105-111, 118-158 and 1-97 of Seq ID No 210; 7-21, 30-48, 51-58, 60-85, 94-123, 134-156, 160-167, 169-183, 186-191, 216-229, 237-251, 257-267, 272-282, 287-298 and 220-243 of Seq ID No 211; 6-29, 34-47, 56-65, 69-76, 83-90, 123-134, 143-151, 158-178, 197-203, 217-235, 243-263, 303-309, 320-333, 338-348, 367-373, 387-393, 407-414, 416-427, 441-457, 473-482, 487-499, 501-509, 514-520, 530-535, 577-583, 590-602, 605-612, 622-629, 641-670, 678-690, 37-71 and 238-307 of Seq ID No 212; 7-40, 121-132, 148-161, 196-202, 209-215, 221-235, 248-255, 271-280, 288-295, 330-339, 395-409, 414-420, 446-451, 475-487, 556-563, 568-575, 580-586, 588-595,

633-638, 643-648, 652-659, 672-685, 695-700, 710-716, 737-742, 749-754, 761-767, 775-781, 796-806, 823-835, 850-863, 884-890, 892-900, 902-915, 934-941 and 406-521 of Seq ID No 213; 9-18, 24-46, 51-58, 67-77, 85-108, 114-126, 129-137, 139-146, 152-165, 173-182, 188-195, 197-204, 217-250, 260-274, 296-313, 343-366, 368-384, 427-434, 437-446, 449-455, 478-484, 492-506, 522-527, 562-591, 599-606, 609-618, 625-631, 645-652 and 577-654 of Seq ID No 214; 13-20, 26-37, 41-53, 56-65, 81-100, 102-114, 118-127, 163-188, 196-202, 231-238, 245-252, 266-285, 293-298, 301-306 and 19-78 of Seq ID No 215; 10-23, 32-42, 54-66, 73-91, 106-113, 118-127, 139-152, 164-173, 198-207, 210-245, 284-300, 313-318, 330-337, 339-346, 354-361, 387-393, 404-426, 429-439, 441-453, 467-473, 479-485, 496-509, 536-544, 551-558, 560-566, 569-574, 578-588, 610-615, 627-635, 649-675, 679-690, 698-716, 722-734, 743-754, 769-780, 782-787 and 480-550 of Seq ID No 216; 6-39, 42-50, 60-68, 76-83, 114-129, 147-162, 170-189, 197-205, 217-231, 239-248, 299-305, 338-344, 352-357, 371-377, 380-451, 459-483, 491-499, 507-523, 537-559, 587-613, 625-681, 689-729, 737-781, 785-809, 817-865, 873-881, 889-939, 951-975, 983-1027, 1031-1055, 1063-1071, 1079-1099, 1103-1127, 1151-1185, 1197-1261, 1269-1309, 1317-1333, 1341-1349, 1357-1465, 1469-1513, 1517-1553, 1557-1629, 1637-1669, 1677-1701, 1709-1725, 1733-1795, 1823-1849, 1861-1925, 1933-1973, 1981-2025, 2029-2053, 2061-2109, 2117-2125, 2133-2183, 2195-2219, 2227-2271, 2275-2299, 2307-2315, 2323-2343, 2347-2371, 2395-2429, 2441-2529, 2537-2569, 2577-2601, 2609-2625, 2633-2695, 2699-2737, 2765-2791, 2803-2867, 2889-2913, 2921-2937, 2945-2969, 2977-2985, 2993-3009, 3023-3045, 3073-3099, 3111-3167, 3175-3215, 3223-3267, 3271-3295, 3303-3351, 3359-3367, 3375-3425, 3437-3461, 3469-3513, 3517-3541, 3549-3557, 3565-3585, 3589-3613, 3637-3671, 3683-3747, 3755-3795, 3803-3819, 3827-3835, 3843-3951, 3955-3999, 4003-4039, 4043-4115, 4123-4143, 4147-4171, 4195-4229, 4241-4305, 4313-4353, 4361-4377, 4385-4393, 4401-4509, 4513-4557, 4561-4597, 4601-4718, 4749-4768, 74-171, 452-559 and 2951-3061 of Seq ID No 217; 16-22, 30-51, 70-111, 117-130, 137-150, 171-178, 180-188, 191-196 and 148-181 of Seq ID No 218; 6-19, 21-46, 50-56, 80-86, 118-126, 167-186, 189-205, 211-242, 244-267, 273-286, 290-297, 307-316, 320-341 and 34-60 of Seq ID No 219; 5-26, 33-43, 48-54, 58-63, 78-83, 113-120, 122-128, 143-152, 157-175, 185-192, 211-225, 227-234, 244-256, 270-281, 284-290, 304-310, 330-337, 348-355, 362-379, 384-394, 429-445, 450-474, 483-490, 511-520, 537-546, 548-554, 561-586, 590-604, 613-629, 149-186, 285-431 and 573-659 of Seq ID No 220; 5-26, 49-59, 61-67, 83-91, 102-111, 145-157, 185-192, 267-272, 279-286, 292-298, 306-312, 134-220, 235-251 and 254-280 of Seq ID No 221; 5-19, 72-79, 83-92, 119-124, 140-145, 160-165, 167-182, 224-232, 240-252, 259-270, 301-310, 313-322, 332-343, 347-367, 384-398, 416-429, 431-446, 454-461 and 1-169 of Seq ID No 222; 8-17, 26-31, 56-62, 75-83, 93-103, 125-131, 135-141, 150-194, 205-217, 233-258, 262-268, 281-286 and 127-168 of Seq ID No 223; 6-12, 69-75, 108-115, 139-159, 176-182, 194-214 and 46-161 of Seq ID No 224; 6-13, 18-27, 39-48, 51-59, 66-73, 79-85, 95-101, 109-116, 118-124, 144-164, 166-177, 183-193, 197-204, 215-223, 227-236, 242-249, 252-259, 261-270, 289-301, 318-325 and 12-58 of Seq ID No 225; 4-10, 26-32, 48-60, 97-105, 117-132, 138-163, 169-185, 192-214, 219-231, 249-261, 264-270, 292-308, 343-356, 385-392, 398-404, 408-417, 435-441 and 24-50 of Seq ID No 226; 10-40, 42-48, 51-61, 119-126 and 1-118 of Seq ID No 227; 5-17, 40-58, 71-83, 103-111, 123-140, 167-177, 188-204 and 116-128 of Seq ID No 228; 4-9, 11-50, 57-70, 112-123, 127-138 and 64-107 of Seq ID No 229; 9-39, 51-67 and 1-101 of Seq ID No 230; 5-14, 17-25, 28-46, 52-59, 85-93, 99-104, 111-120, 122-131, 140-148, 158-179, 187-197, 204-225, 271-283, 285-293 and 139-155 of Seq ID No 231; 42-70, 73-90, 92-108, 112-127, 152-164, 166-172, 181-199, 201-210, 219-228, 247-274, 295-302, 322-334, 336-346, 353-358, 396-414, 419-425, 432-438, 462-471, 518-523, 531-536, 561-567, 576-589, 594-612, 620-631, 665-671, 697-710, 718-731, 736-756, 765-771, 784-801 and 626-653 of Seq ID No 232; 8-28, 41-51, 53-62, 68-74, 79-85, 94-100, 102-108, 114-120, 130-154, 156-162, 175-180, 198-204, 206-213, 281-294, 308-318, 321-339, 362-368, 381-386, 393-399, 407-415 and 2-13 of Seq ID No 233; 4-39, 48-65, 93-98, 106-112, 116-129 and 10-36 of Seq ID No 234; 25-32, 35-50, 66-71, 75-86, 90-96, 123-136, 141-151, 160-179, 190-196, 209-215, 222-228, 235-242, 257-263, 270-280 and 209-247 of Seq ID No 235; 5-29, 31-38, 50-57, 62-75, 83-110, 115-132, 168-195, 197-206, 216-242, 249-258, 262-269, 333-340, 342-350, 363-368, 376-392, 400-406, 410-421, 423-430, 436-442, 448-454, 460-466, 471-476, 491-496, 511-516, 531-536, 551-556, 571-576, 585-591, 599-605, 27-70, 219-293, 441-504 and 512-584 of Seq ID No 236; 4-12, 14-34, 47-75, 83-104, 107-115, 133-140, 148-185, 187-196, 207-212, 224-256, 258-265, 281-287, 289-296, 298-308, 325-333, 345-355, 365-371, 382-395, 424-435, 441-457, 465-472, 483-491, 493-505, 528-534, 536-546, 552-558, 575-584, 589-600, 616-623 and 576-591 of Seq ID No 237; 4-76, 78-89, 91-126, 142-148, 151-191, 195-208, 211-223, 226-240, 256-277, 279-285, 290-314, 317-323, 358-377, 381-387, 391-396, 398-411, 415-434, 436-446, 454-484, 494-512, 516-523, 538-552, 559-566, 571-577, 579-596, 599-615, 620-627, 635-644, 694-707, 720-734, 737-759, 761-771 and 313-329 of Seq ID No

238; 7-38, 44-49, 79-89, 99-108, 117-123, 125-132, 137-146, 178-187, 207-237, 245-255, 322-337, 365-387, 398-408, 445-462, 603-608, 623-628, 644-650, 657-671, 673-679 and 111-566 of Seq ID No 239; 6-20, 22-35, 39-45, 58-64, 77-117, 137-144, 158-163, 205-210, 218-224, 229-236, 239-251, 263-277, 299-307, 323-334, 353-384, 388-396, 399-438, 443-448, 458-463, 467-478, 481-495, 503-509, 511-526, 559-576, 595-600, 612-645, 711-721, 723-738, 744-758, 778-807 and 686-720 of Seq ID No 240; 10-33, 35-41, 72-84, 129-138, 158-163, 203-226, 243-252, 258-264, 279-302, 322-329, 381-386, 401-406, 414-435 and 184-385 of Seq ID No 241; 4-9, 19-24, 41-47, 75-85, 105-110, 113-146 and 45-62 of Seq ID No 242; 4-25, 52-67, 117-124, 131-146, 173-180, 182-191, 195-206, 215-221, 229-236, 245-252, 258-279, 286-291, 293-302, 314-320, 327-336, 341-353, 355-361, 383-389 and 1-285 of Seq ID No 243; 14-32, 38-50, 73-84, 93-105, 109-114 and 40-70 of Seq ID No 244; 5-26 and 22-34 of Seq ID No 245; 23-28 and 13-39 of Seq ID No 246; 8-14 and 21-34 of Seq ID No 247; 4-13, 20-29, 44-50, 59-74 and 41-69 of Seq ID No 248; 4-9, 19-42, 48-59, 71-83 and 57-91 of Seq ID No 249; 4-14 and 10-28 of Seq ID No 250; 22-28, 32-42, 63-71, 81-111, 149-156, 158-167, 172-180, 182-203, 219-229 and 27-49 of Seq ID No 251; 17-27 and 23-32 of Seq ID No 252; 18-24 and 28-38 of Seq ID No 253; 9-15 and 13-27 of Seq ID No 254; 13-22 and 18-29 of Seq ID No 255; 17-26 and 2-11 of Seq ID No 256; 4-33 and 16-32 of Seq ID No 257; 4-10, 37-43, 54-84, 92-127 and 15-62 of Seq ID No 258; 4-14, 20-32, 35-60, 69-75, 79-99, 101-109, 116-140 and 124-136 of Seq ID No 259; 2-13 of Seq ID No 260; 4-13, 28-42 and 42-57 of Seq ID No 261; 4-14, 27-44 and 14-35 of Seq ID No 262; 4-12 and 1-27 of Seq ID No 263; 4-18, 39-45, 47-74 and 35-66 of Seq ID No 264; 8-20, 43-77 and 17-36 of Seq ID No 265; 4-30, 35-45, 51-57 and 35-49 of Seq ID No 266; 4-24, 49-57 and 15-34 of Seq ID No 267; 4-22 and 8-27 of Seq ID No 268; 13-25, 32-59, 66-80 and 21-55 of Seq ID No 269; 4-10, 24-33, 35-42, 54-65, 72-82, 98-108 and 15-30 of Seq ID No 270; 8-19 and 17-47 of Seq ID No 271; 12-18, 40-46 and 31-52 of Seq ID No 272; 4-20, 35-78, 83-102, 109-122 and 74-86 of Seq ID No 273; 7-17, 21-41, 46-63 and 2-20 of Seq ID No 274; 30-37 and 2-33 of Seq ID No 275; 4-13, 17-25 and 1-15 of Seq ID No 276; 17-31, 44-51 and 20-51 of Seq ID No 277; 20-30 and 5-23 of Seq ID No 278; 13-33, 48-71 and 92-110 of Seq ID No 279; 4-9, 50-69, 76-88, 96-106, 113-118 and 12-34 of Seq ID No 280; 4-24 and 6-26 of Seq ID No 281; 7-26 and 14-30 of Seq ID No 282; 9-39, 46-68, 75-82, 84-103 and 26-44 of Seq ID No 283; 4-30, 33-107 and 58-84 of Seq ID No 284; 4-12 and 9-51 of Seq ID No 285; 12-18, 29-37 and 6-37 of Seq ID No 286; 4-21, 33-52, 64-71 and 16-37 of Seq ID No 287; 9-19 and 2-30 of Seq ID No 288; 20-37 of Seq ID No 245; 8-27 of Seq ID No 246; 10-27 of Seq ID No 247; 42-59 and 52-69 of Seq ID No 248; 63-80 and 74-91 of Seq ID No 249; 11-28 of Seq ID No 250; 28-49 of Seq ID No 251; 15-32 of Seq ID No 252; 4-20 of Seq ID No 253; 10-27 of Seq ID No 254; 17-34 of Seq ID No 255; 1-18 of Seq ID No 256; 16-33 of Seq ID No 257; 16-36, 30-49 and 43-62 of Seq ID No 258; 122-139 of Seq ID No 259; 1-18 of Seq ID No 260; 41-58 of Seq ID No 261; 15-35 of Seq ID No 262; 2-27 of Seq ID No 263; 18-36 of Seq ID No 265; 34-51 of Seq ID No 266; 9-27 of Seq ID No 268; 22-47 of Seq ID No 269; 18-36 and 29-47 of Seq ID No 271; 32-52 of Seq ID No 272; 72-89 of Seq ID No 273; 3-20 of Seq ID No 274; 3-21 and 15-33 of Seq ID No 275; 1-18 of Seq ID No 276; 6-23 of Seq ID No 278; 93-110 of Seq ID No 279; 13-34 of Seq ID No 280; 7-26 and 9-26 of Seq ID No 281; 16-33 of Seq ID No 282; 27-44 of Seq ID No 283; 67-84 of Seq ID No 284; 10-33 and 26-50 of Seq ID No 285; 7-25 and 19-37 of Seq ID No 286; 17-37 of Seq ID No 287; 3-20 and 13-30 of Seq ID No 288; 62-80 and 75-93 of Seq ID No 145; 92-108 of Seq ID No 147; 332-349, 177-200 and 1755-1777 of Seq ID No 148; 109-133, 149-174, 260-285 and 460-485 of Seq ID No 149; 26-47 and 42-64 of Seq ID No 150; 22-41, 35-54, 115-130, 306-325, 401-420 and 454-478 of Seq ID No 151; 22-45 of Seq ID No 155; 156-174, 924-940, 1485-1496, 1447-1462 and 1483-1498 of Seq ID No 160; 457-475 of Seq ID No 161; 302-325 of Seq ID No 163; 288-305 of Seq ID No 164; 244-266 and 260-282 of Seq ID No 165; 204-225 and 220-241 of Seq ID No 166; 324-345, 340-361, 356-377, 372-393 and 388-408 of Seq ID No 167; 39-64 of Seq ID No 168; 54-76 and 70-92 of Seq ID No 169; 1227-1247, 1539-1559, 1554-1574, 1569-1589, 1584-1604, 1242-1262, 1272-1292, 1287-1308, 1456-1477, 1472-1494, 1488-1510 and 1505-1526 of Seq ID No 170; 351-368 of Seq ID No 172; 179-200, 195-216, 211-232, 227-248 and 243-263 of Seq ID No 173; 13-37 of Seq ID No 174; 208-224 of Seq ID No 175; 42-64, 59-81, 304-328, 323-348, 465-489, 968-992, 1399-1418, 1412-1431 and 2092-2111 of Seq ID No 177; 1528-1547 and 1541-1560 of Seq ID No 178; 184-200, 367-388, 382-403, 409-429, 425-444 and 438-457 of Seq ID No 179; 27-50 and 45-67 of Seq ID No 180; 114-131 and 405-419 of Seq ID No 183; 113-134, 129-150, 145-166, 161-182 and 177-198 of Seq ID No 184; 495-515 of Seq ID No 186; 346-358 of Seq ID No 187; 208-224 of Seq ID No 190; 178-194, 202-223, 217-238, 288-308 and 1355-1372 of Seq ID No 192; 57-78 of Seq ID No 194; 347-369, 364-386, 381-403, 398-420, 415-437 and 432-452 of Seq ID No 197; 347-372 of Seq ID No 198;

147-163 of Seq ID No 199; 263-288 of Seq ID No 200; 361-377 of Seq ID No 202; 82-104, 99-121, 116-138, 133-155 and 150-171 of Seq ID No 204; 110-130 and 125-145 of Seq ID No 205; 613-631, 626-644 and 196-213 of Seq ID No 206; 78-100, 95-117, 112-134 and 129-151 of Seq ID No 208; 158-180, 175-197, 192-214, 209-231 and 226-248 of Seq ID No 209; 30-50, 45-65 and 60-79 of Seq ID No 210; 431-455 and 450-474 of Seq ID No 213; 579-601, 596-618, 613-635 and 630-653 of Seq ID No 214; 920-927, 98-119, 114-135, 130-151, 146-167 and 162-182 of Seq ID No 217; 36-59 of Seq ID No 219; 194-216 and 381-404 of Seq ID No 220; 236-251 and 255-279 of Seq ID No 221; 80-100 and 141-164 of Seq ID No 222; 128-154 of Seq ID No 223; 82-100, 95-116 and 111-134 of Seq ID No 224; 55-76, 71-92 and 87-110 of Seq ID No 227; 91-106 of Seq ID No 229; 74-96 of Seq ID No 230; 140-157 of Seq ID No 231; 4-13 of Seq ID No 233; 41-65 and 499-523 of Seq ID No 236; 122-146, 191-215, 288-313, 445-469 and 511-535 of Seq ID No 239; 347-368 of Seq ID No 241; 46-61 of Seq ID No 242; 15-37, 32-57, 101-121, 115-135, 138-158, 152-172, 220-242 and 236-258 of Seq ID No 243.

The present invention also provides a process for producing a S. pneumoniae hyperimmune serum reactive antigen or a fragment thereof according to the present invention comprising expressing one or more of the nucleic acid molecules according to the present invention in a suitable expression system.

Moreover, the present invention provides a process for producing a cell, which expresses a S pneumoniae hyperimmune serum reactive antigen or a fragment thereof according to the present invention comprising transforming or transfecting a suitable host cell with the vector according to the present invention.

According to the present invention a pharmaceutical composition, especially a vaccine, comprising a hyperimmune serum-reactive antigen or a fragment thereof as defined in the present invention or a nucleic acid molecule as defined in the present invention is provided.

In a preferred embodiment the pharmaceutical composition further comprises an immunostimulatory substance, preferably selected from the group comprising polycationic polymers, especially polycationic peptides, immunostimulatory deoxynucleotides (ODNs), peptides containing at least two LysLeuLys motifs, especially KLKLLLLLKLK, neuroactive compounds, especially human growth hormone, alumn, Freund's complete or incomplete adjuvants or combinations thereof.

In a more preferred embodiment the immunostimulatory substance is a combination of either a polycationic polymer and immunostimulatory deoxynucleotides or of a peptide containing at least two LysLeuLys motifs and immunostimulatory deoxynucleotides.

In a still more preferred embodiment the polycationic polymer is a polycationic peptide, especially polyarginine.

According to the present invention the use of a nucleic acid molecule according to the present invention or a hyperimmune serum-reactive antigen or fragment thereof according to the present invention for the manufacture of a pharmaceutical preparation, especially for the manufacture of a vaccine against S. pneumoniae infection, is provided.

Also an antibody, or at least an effective part thereof, which binds at least to a selective part of the hyperimmune serum-reactive antigen or a fragment thereof according to the present invention, is provided herewith.

In a preferred embodiment the antibody is a monoclonal antibody.

In another preferred embodiment the effective part of the antibody comprises Fab fragments.

In a further preferred embodiment the antibody is a chimeric antibody.

In a still preferred embodiment the antibody is a humanized antibody.

The present invention also provides a hybridoma cell line, which produces an antibody according to the present invention.

Moreover, the present invention provids a method for producing an antibody according to the present invention, characterized by the following steps:

- initiating an immune response in a non-human animal by administrating an hyperimmune serum-reactive antigen or a fragment thereof, as defined in the invention, to said animal,
- removing an antibody containing body fluid from said animal, and
- producing the antibody by subjecting said antibody containing body fluid to further purification steps.

Accordingly, the present invention also provides a method for producing an antibody according to the present invention, characterized by the following steps:

- initiating an immune response in a non-human animal by administrating an hyperimmune serum-reactive antigen or a fragment thereof, as defined in the present invention, to said animal,
- removing the spleen or spleen cells from said animal,
- producing hybridoma cells of said spleen or spleen cells,
- selecting and cloning hybridoma cells specific for said hyperimmune serum-reactive antigens or a fragment thereof,
- producing the antibody by cultivation of said cloned hybridoma cells and optionally further purification steps.

The antibodies provided or produced according to the above methods may be used for the preparation of a medicament for treating or preventing S. pneumoniae infections.

According to another aspect the present invention provides an antagonist, which binds to a hyperimmune serum-reactive antigen or a fragment thereof according to the present invention.

Such an antagonist capable of binding to a hyperimmune serum-reactive antigen or fragment thereof according to the present invention may be identified by a method comprising the following steps:

a) contacting an isolated or immobilized hyperimmune serum-reactive antigen or a fragment thereof according to the present invention with a candidate antagonist under conditions to permit binding of said candidate antagonist to said hyperimmune serum-reactive antigen or fragment, in the presence of a component capable of providing a detectable signal in response to the binding of the candidate antagonist to said hyperimmune serum reactive antigen or fragment thereof; and

b) detecting the presence or absence of a signal generated in response to the binding of the antagonist to the hyperimmune serum reactive antigen or the fragment thereof.

An antagonist capable of reducing or inhibiting the interaction activity of a hyperimmune serum-reactive antigen or a fragment thereof according to the present invention to its interaction partner may be identified by a method comprising the following steps:

a) providing a hyperimmune serum reactive antigen or a hyperimmune fragment thereof according to the present invention,

b) providing an interaction partner to said hyperimmune serum reactive antigen or a fragment thereof, especially an antibody according to the present invention,

c) allowing interaction of said hyperimmune serum reactive antigen or fragment thereof to said interaction partner to form an interaction complex,

d)   providing a candidate antagonist,

e)   allowing a competition reaction to occur between the candidate antagonist and the interaction complex,

f)   determining whether the candidate antagonist inhibits or reduces the interaction activities of the hyperimmune serum reactive antigen or the fragment thereof with the interaction partner.

The hyperimmune serum reactive antigens or fragments thereof according to the present invention may be used for the isolation and/or purification and/or identification of an interaction partner of said hyperimmune serum reactive antigen or fragment thereof.

The present invention also provides a process for *in vitro* diagnosing a disease related to expression of a hyperimmune serum-reactive antigen or a fragment thereof according to the present invention comprising determining the presence of a nucleic acid sequence encoding said hyperimmune serum reactive antigen or fragment thereof according to the present invention or the presence of the hyperimmune serum reactive antigen or fragment thereof according to the present invention.

The present invention also provides a process for *in vitro* diagnosis of a bacterial infection, especially a *S. pneumoniae* infection, comprising analyzing for the presence of a nucleic acid sequence encoding said hyperimmune serum reactive antigen or fragment thereof according to the present invention or the presence of the hyperimmune serum reactive antigen or fragment thereof according to the present invention.

Moreover, the present invention provides the use of a hyperimmune serum reactive antigen or fragment thereof according to the present invention for the generation of a peptide binding to said hyperimmune serum reactive antigen or fragment thereof, wherein the peptide is an anticaline.

The present invention also provides the use of a hyperimmune serum-reactive antigen or fragment thereof according to the present invention for the manufacture of a functional nucleic acid, wherein the functional nucleic acid is selected from the group comprising aptamers and spiegelmers.

The nucleic acid molecule according to the present invention may also be used for the manufacture of a functional ribonucleic acid, wherein the functional ribonucleic acid is selected from the group comprising ribozymes, antisense nucleic acids and siRNA.

The present invention advantageously provides an efficient, relevant and comprehensive set of isolated nucleic acid molecules and their encoded hyperimmune serum reactive antigens or fragments thereof identified from *S. pneumoniae* using an antibody preparation from multiple human plasma pools and surface expression libraries derived from the genome of *S. pneumoniae*. Thus, the present invention fulfils a widely felt demand for *S. pneumoniae* antigens, vaccines, diagnostics and products useful in procedures for preparing antibodies and for identifying compounds effective against *S. pneumoniae* infection.

An effective vaccine should be composed of proteins or polypeptides, which are expressed by all strains and are able to induce high affinity, abundant antibodies against cell surface components of S. *pneumoniae*. The antibodies should be IgG1 and/or IgG3 for opsonization, and any IgG subtype and IgA for neutralisation of adherence and toxin action. A chemically defined vaccine must be definitely superior compared to a whole cell vaccine (attenuated or killed), since components of S. *pneumoniae*, which cross-react with human tissues or inhibit opsonization can be eliminated, and the individual proteins inducing protective antibodies and/or a protective immune response can be selected.

The approach, which has been employed for the present invention, is based on the interaction of pneumococcal proteins or peptides with the antibodies present in human sera. The antibodies produced against S. *pneumoniae* by the human immune system and present in human sera are indicative of the *in*

vivo expression of the antigenic proteins and their immunogenicity. In addition, the antigenic proteins as identified by the bacterial surface display expression libraries using pools of pre-selected sera, are processed in a second and third round of screening by individual selected or generated sera. Thus the present invention supplies an efficient, relevant, comprehensive set of pneumoococcal antigens as a pharmaceutical composition, especially a vaccine preventing infection by *S. pneumoniae.*

In the antigen identification program for identifying a comprehensive set of antigens according to the present invention, at least two different bacterial surface expression libraries are screened with several serum pools or plasma fractions or other pooled antibody containing body fluids (antibody pools). The antibody pools are derived from a serum collection, which has been tested against antigenic compounds of *S. pneumoniae,* such as whole cell extracts and culture supernatant proteins. Preferably, two distinct serum collections are used: 1. With very stable antibody repertoire: normal adults, clinically healthy people, who are non-carriers and overcame previous encounters or currently carriers of *S. pneumoniae* without acute disease and symptoms, 2. With antibodies induced acutely by the presence of the pathogenic organism: patients with acute disease with different manifestations (e.g. *S. pneumoniae* pharyngitis, pneumonia, bacteraemia, peritonitis, meningitis and sepsis). Sera have to react with multiple Pneumococcus-specific antigens in order to be considered hyperimmune and therefore relevant in the screening method applied for the present invention.

The expression libraries as used in the present invention should allow expression of all potential antigens, e.g. derived from all secreted and surface proteins of *S. pneumoniae.* Bacterial surface display libraries will be represented by a recombinant library of a bacterial host displaying a (total) set of expressed peptide sequences of *S. pneumoniae* on two selected outer membrane proteins (LamB and FhuA) at the bacterial host membrane {Georgiou, G., 1997}; {Etz, H. et al., 2001}. One of the advantages of using recombinant expression libraries is that the identified hyperimmune serum-reactive antigens may be instantly produced by expression of the coding sequences of the screened and selected clones expressing the hyperimmune serum-reactive antigens without further recombinant DNA technology or cloning steps necessary.

The comprehensive set of antigens identified by the described program according to the present invention is analysed further by one or more additional rounds of screening. Therefore individual antibody preparations or antibodies generated against selected peptides, which were identified as immunogenic are used. According to a preferred embodiment the individual antibody preparations for the second round of screening are derived from patients who have suffered from an acute infection with *S. pneumoniae,* especially from patients who show an antibody titer above a certain minimum level, for example an antibody titer being higher than 80 percentile, preferably higher than 90 percentile, especially higher than 95 percentile of the human (patient or healthy individual) sera tested. Using such high titer individual antibody preparations in the second screening round allows a very selective identification of the hyperimmune serum-reactive antigens and fragments thereof from *S. pneumoniae.*

Following the comprehensive screening procedure, the selected antigenic proteins, expressed as recombinant proteins or *in vitro* translated products, in case it can not be expressed in prokaryotic expression systems, or the identified antigenic peptides (produced synthetically) are tested in a second screening by a series of ELISA and Western blotting assays for the assessment of their immunogenicity with a large human serum collection (minimum ~150 healthy and patients sera).

It is important that the individual antibody preparations (which may also be the selected serum) allow a selective identification of the most promising candidates of all the hyperimmune serum-reactive antigens from all the promising candidates from the first round. Therefore, preferably at least 10 individual antibody preparations (i.e. antibody preparations (e.g. sera) from at least 10 different individuals having suffered from an infection to the chosen pathogen) should be used in identifying these antigens in the second screening round. Of course, it is possible to use also less than 10 individual preparations,

- 16 -

however, selectivity of the step may not be optimal with a low number of individual antibody preparations. On the other hand, if a given hyperimmune serum-reactive antigen (or an antigenic fragment thereof) is recognized by at least 10 individual antibody preparations, preferably at least 30, especially at least 50 individual antibody preparations, identification of the hyperimmune serum-reactive antigen is also selective enough for a proper identification. Hyperimmune serum-reactivity may of course be tested with as many individual preparations as possible (e.g. with more than 100 or even with more than 1,000).

Therefore, the relevant portion of the hyperimmune serum-reactive antibody preparations according to the method of the present invention should preferably be at least 10, more preferred at least 30, especially at least 50 individual antibody preparations. Alternatively (or in combination) hyperimmune serum-reactive antigens may preferably be also identified with at least 20%, preferably at least 30%, especially at least 40% of all individual antibody preparations used in the second screening round.

According to a preferred embodiment of the present invention, the sera from which the individual antibody preparations for the second round of screening are prepared (or which are used as antibody preparations), are selected by their titer against S. pneumoniae (e.g. against a preparation of this pathogen, such as a lysate, cell wall components and recombinant proteins). Preferably, some are selected with a total IgA titer above 2,000 U, especially above 4,000 U, and/or an IgG titer above 5,000 U, especially above 12,000 U (U = units, calculated from the OD_{405nm} reading at a given dilution) when the whole organism (total lysate or whole cells) is used as antigen in the ELISA.

The antibodies produced against streptococci by the human immune system and present in human sera are indicative of the *in vivo* expression of the antigenic proteins and their immunogenicity. The recognition of linear epitopes recognized by serum antibodies can be based on sequences as short as 4-5 amino acids. Of course it does not necessarily mean that these short peptides are capable of inducing the given antibody *in vivo*. For that reason the defined epitopes, polypeptides and proteins are further to be tested in animals (mainly in mice) for their capacity to induce antibodies against the selected proteins in vivo.

The preferred antigens are located on the cell surface or are secreted, and are therefore accessible extracellularly. Antibodies against cell wall proteins are expected to serve multiple purposes: to inhibit adhesion, to interfere with nutrient acquisition, to inhibit immune evasion nand to promote phagocytosis (Hornef, M. et al., 2002). Antibodies against secreted proteins are beneficial in neutralisation of their function as toxin or virulence component. It is also known that bacteria communicate with each other through secreted proteins. Neutralizing antibodies against these proteins will interrupt growth-promoting cross-talk between or within streptococcal species. Bioinformatic analyses (signal sequences, cell wall localisation signals, transmembrane domains) proved to be very useful in assessing cell surface localisation or secretion. The experimental approach includes the isolation of antibodies with the corresponding epitopes and proteins from human serum, and the generation of immune sera in mice against (poly)peptides selected by the bacterial surface display screens. These sera are then used in a third round of screening as reagents in the following assays: cell surface staining of S. pneumoniae grown under different conditions (FACS or microscopy), determination of neutralizing capacity (toxin, adherence), and promotion of opsonization and phagocytosis (*in vitro* phagocytosis assay).

For that purpose, bacterial E. coli clones are directly injected into mice and immune sera are taken and tested in the relevant in vitro assay for functional opsonic or neutralizing antibodies. Alternatively, specific antibodies may be purified from human or mouse sera using peptides or proteins as substrate.

Host defence against S. pneumoniae relies mainly on opsonophagocytic killing mechanism. Inducing high affinity antibodies of the opsonic and neutralizing type by vaccination helps the innate immune system to eliminate bacteria and toxins. This makes the method according to the present invention an optimal tool

- 17 -

for the identification of pneumococcal antigenic proteins.

The skin and mucous membranes are formidable barriers against invasion by streptococci. However, once the skin or the mucous membranes are breached the first line of non-adaptive cellular defence begins its co-ordinate action through complement and phagocytes, especially the polymorphonuclear leukocytes (PMNs). These cells can be regarded as the cornerstones in eliminating invading bacteria. As *Streptococcus pneumoniae* is a primarily extracellular pathogen, the major anti-streptococcal adaptive response comes from the humoral arm of the immune system, and is mediated through three major mechanisms: promotion of opsonization, toxin neutralisation, and inhibition of adherence. It is believed that opsonization is especially important, because of its requirement for an effective phagocytosis. For efficient opsonization the microbial surface has to be coated with antibodies and complement factors for recognition by PMNs through receptors to the Fc fragment of the IgG molecule or to activated C3b. After opsonization, streptococci are phagocytosed and killed. Antibodies bound to specific antigens on the cell surface of bacteria serve as ligands for the attachment to PMNs and to promote phagocytosis. The very same antibodies bound to the adhesins and other cell surface proteins are expected to neutralize adhesion and prevent colonization. The selection of antigens as provided by the present invention is thus well suited to identify those that will lead to protection against infection in an animal model or in humans.

According to the antigen identification method used herein, the present invention can surprisingly provide a set of comprehensive novel nucleic acids and novel hyperimmune serum reactive antigens and fragments thereof of *S. pneumoniae*, among other things, as described below. According to one aspect, the invention particularly relates to the nucleotide sequences encoding hyperimmune serum reactive antigens which sequences are set forth in the Sequence listing **Seq ID No:** 1-144, 289-303 and the corresponding encoded amino acid sequences representing hyperimmune serum reactive antigens are set forth in the Sequence Listing **Seq ID No** 145-288 and 304-318.

In a preferred embodiment of the present invention, a nucleic acid molecule is provided which exhibits 70% identity over their entire length to a nucleotide sequence set forth with **Seq ID No 1, 101-144**. Most highly preferred are nucleic acids that comprise a region that is at least 80% or at least 85% identical over their entire length to a nucleic acid molecule set forth with **Seq ID No 1, 101-144**. In this regard, nucleic acid molecules at least 90%, 91%, 92%, 93%, 94%, 95%, or 96% identical over their entire length to the same are particularly preferred. Furthermore, those with at least 97% are highly preferred, those with at least 98% and at least 99% are particularly highly preferred, with at least 99% or 99.5% being the more preferred, with 100% identity being especially preferred. Moreover, preferred embodiments in this respect are nucleic acids which encode hyperimmune serum reactive antigens or fragments thereof (polypeptides) which retain substantially the same biological function or activity as the mature polypeptide encoded by said nucleic acids set forth in the Seq ID No 1, 101-144.

Identity, as known in the art and used herein, is the relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. Identity can be readily calculated. While there exist a number of methods to measure identity between two polynucleotide or two polypeptide sequences, the term is well known to skilled artisans (e.g. *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity are codified in computer programs. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package {Devereux, J. et al., 1984}, BLASTP, BLASTN, and FASTA {Altschul, S. et al., 1990}.

According to another aspect of the invention, nucleic acid molecules are provided which exhibit at least 96% identity to the nucleic acid sequence set forth with **Seq ID No** 2-6, 8, 10-16, 18-23, 25-31, 34, 36, 38-42,

44, 47-48, 51, 53, 55-62, 64, 67, 71-76, 78-79, 81-94, 96-100.

According to a further aspect of the present invention, nucleic acid molecules are provided which are identical to the nucleic acid sequences set forth with **Seq ID No 9, 17, 24, 32, 37, 43, 52, 54, 65-66, 70, 80.**

The nucleic acid molecules according to the present invention can as a second alternative also be a nucleic acid molecule which is at least essentially complementary to the nucleic acid described as the first alternative above. As used herein complementary means that a nucleic acid strand is base pairing via Watson-Crick base pairing with a second nucleic acid strand. Essentially complementary as used herein means that the base pairing is not occurring for all of the bases of the respective strands but leaves a certain number or percentage of the bases unpaired or wrongly paired. The percentage of correctly pairing bases is preferably at least 70 %, more preferably 80 %, even more preferably 90 % and most preferably any percentage higher than 90 %. It is to be noted that a percentage of 70 % matching bases is considered as homology and the hybridization having this extent of matching base pairs is considered as stringent. Hybridization conditions for this kind of stringent hybridization may be taken from Current Protocols in Molecular Biology (John Wiley and Sons, Inc., 1987). More particularly, the hybridization conditions can be as follows:

- Hybridization performed e.g. in 5 x SSPE, 5 x Denhardt's reagent, 0.1% SDS, 100 g/mL sheared DNA at 68°C
- Moderate stringency wash in 0.2xSSC, 0.1% SDS at 42°C
- High stringency wash in 0.1xSSC, 0.1% SDS at 68°C

Genomic DNA with a GC content of 50% has an approximate T_M of 96°C. For 1% mismatch, the T_M is reduced by approximately 1°C.

In addition, any of the further hybridization conditions described herein are in principle applicable as well.

Of course, all nucleic acid sequence molecules which encode the same polypeptide molecule as those identified by the present invention are encompassed by any disclosure of a given coding sequence, since the degeneracy of the genetic code is directly applicable to unambiguously determine all possible nucleic acid molecules which encode a given polypeptide molecule, even if the number of such degenerated nucleic acid molecules may be high. This is also applicable for fragments of a given polypeptide, as long as the fragments encode a polypeptide being suitable to be used in a vaccination connection, e.g. as an active or passive vaccine.

The nucleic acid molecule according to the present invention can as a third alternative also be a nucleic acid which comprises a stretch of at least 15 bases of the nucleic acid molecule according to the first and second alternative of the nucleic acid molecules according to the present invention as outlined above. Preferably, the bases form a contiguous stretch of bases. However, it is also within the scope of the present invention that the stretch consists of two or more moieties, which are separated by a number of bases.

The present nucleic acids may preferably consist of at least 20, even more preferred at least 30, especially at least 50 contiguous bases from the sequences disclosed herein. The suitable length may easily be optimized due to the planned area of use (e.g. as (PCR) primers, probes, capture molecules (e.g. on a (DNA) chip), etc.). Preferred nucleic acid molecules contain at least a contiguous 15 base portion of one or more of the predicted immunogenic amino acid sequences listed in tables 1 and 2, especially the sequences of table 2 with scores of more than 10, preferably more than 20, especially with a score of more than 25. Specifically preferred are nucleic acids containing a contiguous portion of a DNA sequence of any sequence in the sequence protocol of the present application which shows 1 or more, preferably more

than 2, especially more than 5, non-identical nucleic acid residues compared to the published *Streptococcus pneumoniae* strain TIGR4 genome ({Tettelin, H. et al., 2001}; GenBank accession AE005672) and/or any other published *S. pneumoniae* genome sequence or parts thereof, especially of the strain R6 ({Hoskins, J. et al., 2001}; GenBank accession AE007317). Specifically preferred non-identical nucleic acid residues are residues, which lead to a non-identical amino acid residue. Preferably, the nucleic acid sequences encode for polypeptides having at least 1, preferably at least 2, preferably at least 3 different amino acid residues compared to the published *S. pneumoniae* counterparts mentioned above. Also such isolated polypeptides, being fragments of the proteins (or the whole protein) mentioned herein e.g. in the sequence listing, having at least 6, 7, or 8 amino acid residues and being encoded by these nucleic acids are preferred.

The nucleic acid molecule according to the present invention can as a fourth alternative also be a nucleic acid molecule which anneals under stringent hybridisation conditions to any of the nucleic acids of the present invention according to the above outlined first, second, and third alternative. Stringent hybridisation conditions are typically those described herein.

Finally, the nucleic acid molecule according to the present invention can as a fifth alternative also be a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to any of the nucleic acid molecules according to any nucleic acid molecule of the present invention according to the first, second, third, and fourth alternative as outlined above. This kind of nucleic acid molecule refers to the fact that preferably the nucleic acids according to the present invention code for the hyperimmune serum reactive antigens or fragments thereof according to the present invention. This kind of nucleic acid molecule is particularly useful in the detection of a nucleic acid molecule according to the present invention and thus the diagnosis of the respective microorganisms such as *S. pneumoniae* and any disease or diseased condition where this kind of microorganims is involved. Preferably, the hybridisation would occur or be preformed under stringent conditions as described in connection with the fourth alternative described above.

Nucleic acid molecule as used herein generally refers to any ribonucleic acid molecule or deoxyribonucleic acid molecule, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, nucleic acid molecule as used herein refers to, among other, single-and double-stranded DNA, DNA that is a mixture of single- and double-stranded RNA, and RNA that is a mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded, or a mixture of single- and double-stranded regions. In addition, nucleic acid molecule as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term nucleic acid molecule includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "nucleic acid molecule" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are nucleic acid molecule as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term nucleic acid molecule as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of nucleic acid molecule, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, *inter alia*. The term nucleic acid molecule also embraces short nucleic acid molecules often referred to as oligonucleotide(s). "Polynucleotide" and "nucleic acid" or "nucleic acid molecule" are often used interchangeably herein.

Nucleic acid molecules provided in the present invention also encompass numerous unique fragments,

both longer and shorter than the nucleic acid molecule sequences set forth in the sequencing listing of the *S. pneumoniae* coding regions, which can be generated by standard cloning methods. To be unique, a fragment must be of sufficient size to distinguish it from other known nucleic acid sequences, most readily determined by comparing any selected *S. pneumoniae* fragment to the nucleotide sequences in computer databases such as GenBank.

Additionally, modifications can be made to the nucleic acid molecules and polypeptides that are encompassed by the present invention. For example, nucleotide substitutions can be made which do not affect the polypeptide encoded by the nucleic acid, and thus any nucleic acid molecule which encodes a hyperimmune serum reactive antigen or fragments thereof is encompassed by the present invention.

Furthermore, any of the nucleic acid molecules encoding hyperimmune serum reactive antigens or fragments thereof provided by the present invention can be functionally linked, using standard techniques such as standard cloning techniques, to any desired regulatory sequences, whether a *S. pneumoniae* regulatory sequence or a heterologous regulatory sequence, heterologous leader sequence, heterologous marker sequence or a heterologous coding sequence to create a fusion protein.

Nucleic acid molecules of the present invention may be in the form of RNA, such as mRNA or cRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The DNA may be triple-stranded, double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The present invention further relates to variants of the herein above described nucleic acid molecules which encode fragments, analogs and derivatives of the hyperimmune serum reactive antigens and fragments thereof having a deducted *S. pneumoniae* amino acid sequence set forth in the Sequence Listing. A variant of the nucleic acid molecule may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the nucleic acid molecule may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned nucleic acid molecules by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions. Preferred are nucleic acid molecules encoding a variant, analog, derivative or fragment, or a variant, analogue or derivative of a fragment, which have a *S. pneumoniae* sequence as set forth in the Sequence Listing, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid(s) is substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the *S. pneumoniae* polypeptides set forth in the Sequence Listing. Also especially preferred in this regard are conservative substitutions.

The peptides and fragments according to the present invention also include modified epitopes wherein preferably one or two of the amino acids of a given epitope are modified or replaced according to the rules disclosed in e.g. (Tourdot, S. et al., 2000), as well as the nucleic acid sequences encoding such modified epitopes.

It is clear that also epitopes derived from the present epitopes by amino acid exchanges improving, conserving or at least not significantly impeding the T cell activating capability of the epitopes are covered by the epitopes according to the present invention. Therefore the present epitopes also cover epitopes, which do not contain the original sequence as derived from *S. pneumoniae*, but trigger the same or preferably an improved T cell response. These epitope are referred to as "heteroclitic"; they need to

have a similar or preferably greater affinity to MHC/HLA molecules, and the need the ability to stimulate the T cell receptors (TCR) directed to the original epitope in a similar or preferably stronger manner.

Heteroclitic epitopes can be obtained by rational design i.e. taking into account the contribution of individual residues to binding to MHC/HLA as for instance described by {Rammensee, H. et al., 1999}, combined with a systematic exchange of residues potentially interacting with the TCR and testing the resulting sequences with T cells directed against the original epitope. Such a design is possible for a skilled man in the art without much experimentation.

Another possibility includes the screening of peptide libraries with T cells directed against the original epitope. A preferred way is the positional scanning of synthetic peptide libraries. Such approaches have been described in detail for instance by {Hemmer, B. et al., 1999}and the references given therein.

As an alternative to epitopes represented by the present derived amino acid sequences or heteroclitic epitopes, also substances mimicking these epitopes e.g. "peptidemimetica" or "retro-inverso-peptides" can be applied.

Another aspect of the design of improved epitopes is their formulation or modification with substances increasing their capacity to stimulate T cells. These include T helper cell epitopes, lipids or liposomes or preferred modifications as described in WO 01/78767.

Another way to increase the T cell stimulating capacity of epitopes is their formulation with immune stimulating substances for instance cytokines or chemokines like interleukin-2, -7, -12, -18, class I and II interferons (IFN), especially IFN-gamma, GM-CSF, TNF-alpha, flt3-ligand and others.

As discussed additionally herein regarding nucleic acid molecule assays of the invention, for instance, nucleic acid molecules of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding polypeptides of the present invention and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the nucleic acid molecules of the present invention. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 20, at least 25 or at least 30 bases, and may have at least 50 bases. Particularly preferred probes will have at least 30 bases, and will have 50 bases or less, such as 30, 35, 40, 45, or 50 bases.

For example, the coding region of a nucleic acid molecule of the present invention may be isolated by screening a relevant library using the known DNA sequence to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine to which members of the library the probe hybridizes.

The nucleic acid molecules and polypeptides of the present invention may be employed as reagents and materials for development of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to nucleic acid molecule assays, *inter alia*.

The nucleic acid molecules of the present invention that are oligonucleotides can be used in the processes herein as described, but preferably for PCR, to determine whether or not the *S. pneumoniae* genes identified herein in whole or in part are present and/or transcribed in infected tissue such as blood. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained. For this and other purposes the arrays comprising at least one of the nucleic acids according to the present invention as described herein, may be used.

- 22 -

The nucleic acid molecules according to the present invention may be used for the detection of nucleic acid molecules and organisms or samples containing these nucleic acids. Preferably such detection is for diagnosis, more preferable for the diagnosis of a disease related or linked to the present or abundance of *S. pneumoniae.*

Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with *S. pneumoniae* may be identifiable by detecting any of the nucleic acid molecules according to the present invention detected at the DNA level by a variety of techniques. Preferred nucleic acid molecules candidates for distinguishing a *S. pneumoniae* from other organisms can be obtained.

The invention provides a process for diagnosing disease, arising from infection with *S. pneumoniae,* comprising determining from a sample isolated or derived from an individual an increased level of expression of a nucleic acid molecule having the sequence of a nucleic acid molecule set forth in the Sequence Listing. Expression of nucleic acid molecules can be measured using any one of the methods well known in the art for the quantitation of nucleic acid molecules, such as, for example, PCR, RT-PCR, Rnase protection, Northern blotting, other hybridisation methods and the arrays described herein.

Isolated as used herein means separated "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring nucleic acid molecule or a polypeptide naturally present in a living organism in its natural state is not "isolated," but the same nucleic acid molecule or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. As part of or following isolation, such nucleic acid molecules can be joined to other nucleic acid molecules, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated nucleic acid molecules, alone or joined to other nucleic acid molecules such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the nucleic acid molecules and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of nucleic acid molecules or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated nucleic acid molecules or polypeptides within the meaning of that term as it is employed herein.

The nucleic acids according to the present invention may be chemically synthesized. Alternatively, the nucleic acids can be isolated from *S. pneumoniae* by methods known to the one skilled in the art.

According to another aspect of the present invention, a comprehensive set of novel hyperimmune serum reactive antigens and fragments thereof are provided by using the herein described antigen identification method. In a preferred embodiment of the invention, a hyperimmune serum-reactive antigen comprising an amino acid sequence being encoded by any one of the nucleic acids molecules herein described and fragments thereof are provided. In another preferred embodiment of the invention a novel set of hyperimmune serum-reactive antigens which comprises amino acid sequences selected from a group consisting of the polypeptide sequences as represented in Seq ID No 145, 245-288 and fragments thereof are provided. In a further preferred embodiment of the invention hyperimmune serum-reactive antigens which comprise amino acid sequences selected from a group consisting of the polypeptide sequences as represented in Seq ID No 146-150, 152, 154-160, 162-167, 169-175, 178, 180, 182-186, 188, 191-192, 195, 197, 199-206, 208, 211, 215-220, 222-223, 225-238, 240-244 and fragments thereof are provided. In a still preferred embodiment of the invention hyperimmune serum-reactive antigens which comprise amino acid sequences selected from a group consisting of the polypeptide sequences as represented in Seq ID No 153, 161, 168, 176, 181, 187, 196, 198, 209-210, 214, 224 and fragments thereof are provided.

The hyperimmune serum reactive antigens and fragments thereof as provided in the invention include

any polypeptide set forth in the Sequence Listing as well as polypeptides which have at least 70% identity to a polypeptide set forth in the Sequence Listing, preferably at least 80% or 85% identity to a polypeptide set forth in the Sequence Listing, and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide set forth in the Sequence Listing and still more preferably at least 95%, 96%, 97%, 98%, 99% or 99.5% similarity (still more preferably at least 95%, 96%, 97%, 98%, 99%, or 99.5% identity) to a polypeptide set forth in the Sequence Listing and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 4 amino acids and more preferably at least 8, still more preferably at least 30, still more preferably at least 50 amino acids, such as 4, 8, 10, 20, 30, 35, 40, 45 or 50 amino acids.

The invention also relates to fragments, analogs, and derivatives of these hyperimmune serum reactive antigens and fragments thereof. The terms "fragment", "derivative" and "analog" when referring to an antigen whose amino acid sequence is set forth in the Sequence Listing, means a polypeptide which retains essentially the same or a similar biological function or activity as such hyperimmune serum reactive antigen and fragment thereof.

The fragment, derivative or analog of a hyperimmune serum reactive antigen and fragment thereof may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid residues includes a substituent group, or 3) one in which the mature hyperimmune serum reactive antigen or fragment thereof is fused with another compound, such as a compound to increase the half-life of the hyperimmune serum reactive antigen and fragment thereof (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mature hyperimmune serum reactive antigen or fragment thereof, such as a leader or secretory sequence or a sequence which is employed for purification of the mature hyperimmune serum reactive antigen or fragment thereof or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The present invention also relates to antigens of different *S. pneumoniae* isolates. Such homologues may easily be isolated based on the nucleic acid and amino acid sequences disclosed herein. There are more than 90 serotypes in more than 40 serogroups distinguished to date and the typing is based on serotype specific antisera. The presence of any antigen can accordingly be determined for every serotype. In addition it is possible to determine the variability of a particular antigen in the various serotypes as described for the *S. pyogenes* sic gene {Hoe, N. et al., 2001}. The contribution of the various serotypes to the different pneumococcal infections varies in the different age groups and geographical regions {Gray, B. et al., 1979}; {Gray, B. et al., 1986}; {Orange, M. et al., 1993}, reviewed in Epidemiology and Prevention of Vaccine-Preventable Diseases, 7th Edition-Second Printing, The Pink Book). It is an important aspect that the most valuable protective antigens are expected to be conserved among various clinical strains.

Among the particularly preferred embodiments of the invention in this regard are the hyperimmune serum reactive antigens set forth in the Sequence Listing, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of fragments. Additionally, fusion polypeptides comprising such hyperimmune serum reactive antigens, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments are also encompassed by the present invention. Such fusion polypeptides and proteins, as well as nucleic acid molecules encoding them, can readily be made using standard techniques, including standard recombinant techniques for producing and expression of a recombinant polynucleic acid encoding a fusion protein.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another,

among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, having the amino acid sequence of any polypeptide set forth in the Sequence Listing, in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the polypeptide of the present invention. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polypeptides having an amino acid sequence set forth in the Sequence Listing without substitutions.

The hyperimmune serum reactive antigens and fragments thereof of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

Also among preferred embodiments of the present invention are polypeptides comprising fragments of the polypeptides having the amino acid sequence set forth in the Sequence Listing, and fragments of variants and derivatives of the polypeptides set forth in the Sequence Listing.

In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the afore mentioned hyperimmune serum reactive antigen and fragment thereof, and variants or derivative, analogs, fragments thereof. Such fragments may be "free-standing", i.e., not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. Also preferred in this aspect of the invention are fragments characterised by structural or functional attributes of the polypeptide of the present invention, i.e. fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta-amphipathic regions, flexible regions, surface-forming regions, substrate binding regions, and high antigenic index regions of the polypeptide of the present invention, and combinations of such fragments. Preferred regions are those that mediate activities of the hyperimmune serum reactive antigens and fragments thereof of the present invention. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of the hyperimmune serum reactive antigen and fragments thereof of the present invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *S. pneumoniae* or the ability to cause disease in humans. Further preferred polypeptide fragments are those that comprise or contain antigenic or immunogenic determinants in an animal, especially in a human.

An antigenic fragment is defined as a fragment of the identified antigen, which is for itself antigenic or may be made antigenic when provided as a hapten. Therefore, also antigens or antigenic fragments showing one or (for longer fragments) only a few amino acid exchanges are enabled with the present invention, provided that the antigenic capacities of such fragments with amino acid exchanges are not severely deteriorated on the exchange(s), i.e., suited for eliciting an appropriate immune response in an individual vaccinated with this antigen and identified by individual antibody preparations from individual sera.

Preferred examples of such fragments of a hyperimmune serum-reactive antigen are selected from the group consisting of peptides comprising amino acid sequences of column "predicted immunogenic aa", and "Location of identified immunogenic region" of Table 1; the serum reactive epitopes of Table 2, especially peptides comprising amino acid 4-11, 35-64, 66-76, 101-108, 111-119 and 57-114 of Seq ID No

145; 5-27, 32-64, 92-102, 107-113, 119-125, 133-139, 148-162, 177-187, 195-201, 207-214, 241-251, 254-269, 285-300, 302-309, 317-324, 332-357, 365-404, 411-425, 443-463, 470-477, 479-487, 506-512, 515-520, 532-547, 556-596, 603-610, 616-622, 624-629, 636-642, 646-665, 667-674, 687-692, 708-720, 734-739, 752-757, 798-820, 824-851, 856-865 and 732-763 of Seq ID No 146; 14-21, 36-44, 49-66, 102-127, 162-167, 177-196, 45-109 and 145-172 of Seq ID No 147; 17-35, 64-75, 81-92, 100-119, 125-172, 174-183, 214-222, 230-236, 273-282, 287-303, 310-315, 331-340, 392-398, 412-420, 480-505, 515-523, 525-546, 553-575, 592-598, 603-609, 617-625, 631-639, 644-651, 658-670, 681-687, 691-704, 709-716, 731-736, 739-744, 750-763, 774-780, 784-791, 799-805, 809-822, 859-870, 880-885, 907-916, 924-941, 943-949, 973-986, 1010-1016, 1026-1036, 1045-1054, 1057-1062, 1082-1088, 1095-1102, 1109-1120, 1127-1134, 1140-1146, 1152-1159, 1169-1179, 1187-1196; 1243-1251, 1262-1273, 1279-1292, 1306-1312, 1332-1343, 1348-1364, 1379-1390, 1412-1420, 1427-1436, 1458-1468, 1483-1503, 1524-1549, 1574-1588, 1614-1619, 1672-1685, 1697-1707, 1711-1720, 1738-1753, 1781-1787, 1796-1801, 1826-1843, 132-478, 508-592 and 1753-1810 of Seq ID No 148; 15-43, 49-55, 71-77, 104-110, 123-130, 162-171, 180-192, 199-205, 219-227, 246-254, 264-270, 279-287, 293-308, 312-322, 330-342, 349-356, 369-377, 384-394, 401-406, 416-422, 432-439, 450-460, 464-474, 482-494, 501-508, 521-529, 536-546, 553-558, 568-574, 584-591, 602-612, 616-626, 634-646, 653-660, 673-681, 688-698, 705-710, 720-726, 736-749, 833-848, 1-199, 200-337, 418-494 and 549-647 of Seq ID No 149; 9-30, 65-96, 99-123, 170-178 and 1-128 of Seq ID No 150; 7-32, 34-41, 96-106, 127-136, 154-163, 188-199, 207-238, 272-279, 306-312, 318-325, 341-347, 353-360, 387-393, 399-406, 434-440, 452-503, 575-580, 589-601, 615-620, 635-640, 654-660, 674-680, 696-701, 710-731, 1-548 and 660-691 of Seq ID No 151; 4-19, 35-44, 48-59, 77-87, 93-99, 106-111, 130-138, 146-161 and 78-84 of Seq ID No 152; 24-30, 36-43, 64-86, 93-99, 106-130, 132-145, 148-165, 171-177, 189-220, 230-249, 251-263, 293-300, 302-312, 323-329, 338-356, 369-379, 390-412 and 179-193 of Seq ID No 153; 30-39, 61-67, 74-81, 90-120, 123-145, 154-167, 169-179, 182-197, 200-206, 238-244, 267-272 and 230-265 of Seq ID No 154; 14-20, 49-65, 77-86 and 2-68 of Seq ID No 155; 4-9, 26-35, 42-48, 53-61, 63-85, 90-101, 105-111, 113-121, 129-137, 140-150, 179-188, 199-226, 228-237, 248-255, 259-285, 299-308, 314-331, 337-343, 353-364, 410-421, 436-442 and 110-144 of Seq ID No 156; 36-47, 55-63, 94-108, 129-134, 144-158, 173-187, 196-206, 209-238, 251-266, 270-285, 290-295, 300-306, 333-344, 346-354, 366-397, 404-410, 422-435, 439-453, 466-473, 515-523, 529-543, 554-569, 571-585, 590-596, 607-618, 627-643, 690-696, 704-714, 720-728, 741-749, 752-767, 780-799, 225-247 and 480-507 of Seq ID No 157; 16-25, 36-70, 80-93, 100-106 and 78-130 of Seq ID No 158; 18-27, 41-46, 50-57, 65-71, 79-85, 93-98, 113-128, 144-155, 166-178, 181-188, 201-207, 242-262, 265-273, 281-295, 303-309, 318-327 and 36-64 of Seq ID No 159; 7-29, 31-44, 50-59, 91-96, 146-153, 194-201, 207-212, 232-238, 264-278, 284-290, 296-302, 326-353, 360-370, 378-384, 400-405, 409-418, 420-435, 442-460, 499-506, 529-534, 556-562, 564-576, 644-651, 677-684, 687-698, 736-743, 759-766, 778-784, 808-814, 852-858, 874-896, 920-925, 929-935, 957-965, 1003-1012, 1021-1027, 1030-1044, 1081-1087, 1101-1111, 1116-1124, 1148-1159, 1188-1196, 1235-1251, 1288-1303, 1313-1319, 1328-1335, 1367-1373, 1431-1437, 1451-1458, 1479-1503, 1514-1521, 1530-1540, 1545-1552, 1561-1568, 1598-1605, 1617-1647, 1658-1665, 1670-1676, 1679-1689, 1698-1704, 1707-1713, 1732-1738, 1744-1764, 1-70, 154-189, 922-941, 1445-1462 and 1483-1496 of Seq ID No 160; 6-51, 81-91, 104-113, 126-137, 150-159, 164-174, 197-209, 215-224, 229-235, 256-269, 276-282, 307-313, 317-348, 351-357, 376-397, 418-437, 454-464, 485-490, 498-509, 547-555, 574-586, 602-619 and 452-530 of Seq ID No 161; 25-31, 39-47, 49-56, 99-114, 121-127, 159-186, 228-240, 253-269, 271-279, 303-315, 365-382, 395-405, 414-425, 438-453 and 289-384 of Seq ID No 162; 9-24, 41-47, 49-54, 68-78, 108-114, 117-122, 132-140, 164-169, 179-186, 193-199, 206-213, 244-251, 267-274, 289-294, 309-314, 327-333, 209-249 and 286-336 of Seq ID No 163; 9-28, 53-67, 69-82, 87-93, 109-117, 172-177, 201-207, 220-227, 242-247, 262-268, 305-318, 320-325 and 286-306 of Seq ID No 164; 4-10, 26-39, 47-58, 63-73, 86-96, 98-108, 115-123, 137-143, 148-155, 160-176, 184-189, 194-204, 235-240, 254-259, 272-278 and 199-283 of Seq ID No 165; 4-26, 33-39, 47-53, 59-65, 76-83, 91-97, 104-112, 118-137, 155-160, 167-174, 198-207, 242-268, 273-279, 292-315, 320-332, 345-354, 358-367, 377-394, 403-410, 424-439, 445-451, 453-497, 511-518, 535-570, 573-589, 592-601, 604-610 and 202-242 of Seq ID No 166; 8-30, 36-45, 64-71, 76-82, 97-103, 105-112, 134-151, 161-183, 211-234, 253-268, 270-276, 278-284, 297-305, 309-315, 357-362, 366-372, 375-384, 401-407, 409-416, 441-455, 463-470, 475-480, 490-497, 501-513, 524-537, 552-559, 565-576, 581-590, 592-600, 619-625, 636-644, 646-656 and 316-419 of Seq ID No 167; 4-17, 52-58, 84-99, 102-110, 114-120, 124-135, 143-158, 160-173, 177-196, 201-216, 223-250, 259-267, 269-275 and 1-67 of Seq ID No 168; 6-46, 57-67, 69-80, 82-133, 137-143, 147-168, 182-187, 203-209, 214-229, 233-242, 246-280 and 53-93 of Seq ID No 169; 7-40, 50-56, 81-89, 117-123, 202-209, 213-218, 223-229, 248-261, 264-276, 281-288, 303-308, 313-324, 326-332, 340-346, 353-372, 434-443,

465-474, 514-523, 556-564, 605-616, 620-626, 631-636, 667-683, 685-699, 710-719, 726-732, 751-756, 760-771, 779-788, 815-828, 855-867, 869-879, 897-902, 917-924, 926-931, 936-942, 981-1000, 1006-1015, 1017-1028, 1030-1039, 1046-1054, 1060-1066, 1083-1092, 1099-1112, 1122-1130, 1132-1140, 1148-1158, 1161-1171, 1174-1181, 1209-1230, 1236-1244, 1248-1254, 1256-1267, 1269-1276, 1294-1299, 1316-1328, 1332-1354, 1359-1372, 1374-1380, 1384-1390, 1395-1408, 1419-1425, 1434-1446, 1453-1460, 1465-1471, 1474-1493, 1505-1515, 1523-1537, 1547-1555, 1560-1567, 1577-1605, 1633-1651, 1226-1309, 1455-1536 and 1538-1605 of Seq ID No 170; 4-10, 31-39, 81-88, 106-112, 122-135, 152-158, 177-184, 191-197, 221-227, 230-246, 249-255, 303-311, 317-326, 337-344, 346-362, 365-371, 430-437, 439-446, 453-462, 474-484 and 449-467 of Seq ID No 171; 9-15, 24-35, 47-55, 122-128, 160-177, 188-196, 202-208, 216-228, 250-261, 272-303, 318-324, 327-339, 346-352, 355-361, 368-373, 108-218 and 344-376 of Seq ID No 172; 6-14, 17-48, 55-63, 71-90, 99-109, 116-124, 181-189, 212-223, 232-268, 270-294, 297-304, 319-325, 340-348, 351-370, 372-378, 388-394, 406-415, 421-434 and 177-277 of Seq ID No 173; 21-39, 42-61, 65-75, 79-85, 108-115 and 11-38 of Seq ID No 174; 4-17, 26-39, 61-76, 103-113, 115-122, 136-142, 158-192, 197-203, 208-214, 225-230, 237-251 and 207-225 of Seq ID No 175; 5-11, 27-36, 42-53, 62-70, 74-93, 95-104, 114-119, 127-150, 153-159, 173-179, 184-193, 199-206, 222-241, 248-253, 257-280, 289-295, 313-319, 322-342, 349-365, 368-389, 393-406, 408-413, 426-438, 447-461, 463-470, 476-495, 532-537, 543-550 and 225-246 of Seq ID No 176; 4-29, 68-82, 123-130, 141-147, 149-157, 178-191, 203-215, 269-277, 300-307, 327-335, 359-370, 374-380, 382-388, 393-400, 410-417, 434-442, 483-492, 497-503, 505-513, 533-540, 564-569, 601-607, 639-647, 655-666, 693-706, 712-718, 726-736, 752-758, 763-771, 774-780, 786-799, 806-812, 820-828, 852-863, 884-892, 901-909, 925-932, 943-948, 990-996, 1030-1036, 1051-1059, 1062-1068, 1079-1086, 1105-1113, 1152-1162, 1168-1179, 1183-1191, 1204-1210, 1234-1244, 1286-1295, 1318-1326, 1396-1401, 1451-1460, 1465-1474, 1477-1483, 1488-1494, 1505-1510, 1514-1521, 1552-1565, 1593-1614, 1664-1672, 1677-1685, 1701-1711, 1734-1745, 1758-1770, 1784-1798, 1840-1847, 1852-1873, 1885-1891, 1906-1911, 1931-1939, 1957-1970, 1977-1992, 2014-2020, 2026-2032, 2116-2134, 1-348, 373-490, 573-767, 903-1043, 1155-1198, 1243-1482, 1550-1595, 1682-1719, 1793-1921 and 2008-2110 of Seq ID No 177; 10-35, 39-52, 107-112, 181-188, 226-236, 238-253, 258-268, 275-284, 296-310, 326-338, 345-368, 380-389, 391-408, 410-418, 420-429, 444-456, 489-505, 573-588, 616-623, 637-643, 726-739, 741-767, 785-791, 793-803, 830-847, 867-881, 886-922, 949-956, 961-980, 988-1004, 1009-1018, 1027-1042, 1051-1069, 1076-1089, 1108-1115, 1123-1135, 1140-1151, 1164-1179, 1182-1191, 1210-1221, 1223-1234, 1242-1250, 1255-1267, 1281-1292, 1301-1307, 1315-1340, 1348-1355, 1366-1373, 1381-1413, 1417-1428, 1437-1444, 1453-1463, 1478-1484, 1490-1496, 1498-1503, 1520-1536, 1538-1546, 1548-1570, 1593-1603, 1612-1625, 1635-1649, 1654-1660, 1670-1687, 1693-1700, 1705-1711, 1718-1726, 1729-1763, 1790-1813, 1871-1881, 1893-1900, 1907-1935, 1962-1970, 1992-2000, 2006-2013, 2033-2039, 2045-2051, 2055-2067, 2070-2095, 2097-2110, 2115-2121, 2150-2171, 2174-2180, 2197-2202, 2206-2228 and 1526-1560 of Seq ID No 178; 4-17, 35-48, 54-76, 78-107, 109-115, 118-127, 134-140, 145-156, 169-174, 217-226, 232-240, 256-262, 267-273, 316-328, 340-346, 353-360, 402-409, 416-439, 448-456, 506-531, 540-546, 570-578, 586-593, 595-600, 623-632, 662-667, 674-681, 689-705, 713-724, 730-740, 757-763, 773-778, 783-796, 829-835, 861-871, 888-899, 907-939, 941-955, 957-969, 986-1000, 1022-1028, 1036-1044, 1068-1084, 1095-1102, 1118-1124, 1140-1146, 1148-1154, 1168-1181, 1185-1190, 1197-1207, 1218-1226, 1250-1270, 1272-1281, 1284-1296, 1312-1319, 1351-1358, 1383-1409, 1422-1428, 1438-1447, 1449-1461, 1482-1489, 1504-1510, 1518-1527, 1529-1537, 1544-1551, 1569-1575, 1622-1628, 1631-1637, 1682-1689, 1711-1718, 1733-1740, 1772-1783, 1818-1834, 1859-1872, 1-64 and 128-495 of Seq ID No 179; 8-28, 32-37, 62-69, 119-125, 137-149, 159-164, 173-189, 200-205, 221-229, 240-245, 258-265, 268-276, 287-293, 296-302, 323-329 and 1-95 of Seq ID No 180; 9-18, 25-38, 49-63, 65-72, 74-81, 94-117, 131-137, 139-146, 149-158, 162-188, 191-207, 217-225, 237-252, 255-269, 281-293, 301-326, 332-342, 347-354, 363-370, 373-380, 391-400, 415-424, 441-447 and 75-107 of Seq ID No 181; 4-24, 64-71, 81-87, 96-116, 121-128, 130-139, 148-155, 166-173, 176-184, 203-215, 231-238, 243-248, 256-261, 280-286, 288-306, 314-329 and 67-148 of Seq ID No 182; 4-10, 19-37, 46-52, 62-81, 83-89, 115-120, 134-139, 141-151, 168-186, 197-205, 209-234, 241-252, 322-335, 339-345, 363-379, 385-393, 403-431, 434-442, 447-454, 459-465, 479-484, 487-496 and 404-420 of Seq ID No 183; 10-35, 46-66, 71-77, 84-93, 96-122, 138-148, 154-172, 182-213, 221-233, 245-263, 269-275, 295-301, 303-309, 311-320, 324-336, 340-348, 351-359, 375-381 and 111-198 of Seq ID No 184; 14-25, 30-42, 47-61, 67-75, 81-91, 98-106, 114-122, 124-135, 148-193, 209-227 and 198-213 of Seq ID No 185; 5-18, 45-50, 82-90, 97-114, 116-136, 153-161, 163-171, 212-219, 221-227, 240-249, 267-281, 311-317, 328-337, 375-381, 390-395, 430-436, 449-455, 484-495, 538-543, 548-554, 556-564, 580-586, 596-602 and 493-606 of Seq ID No 186; 9-25, 28-34, 37-44, 61-68, 75-81, 88-96, 98-111, 119-133, 138-150, 152-163, 168-182, 186-194, 200-

205, 216-223, 236-245, 257-264, 279-287, 293-304, 311-318, 325-330, 340-346, 353-358, 365-379, 399-409, 444-453 and 303-391 of Seq ID No 187; 16-36, 55-61, 66-76, 78-102, 121-130, 134-146, 150-212, 221-239, 255-276, 289-322, 329-357 and 29-59 of Seq ID No 188; 8-27, 68-74, 77-99, 110-116, 124-141, 171-177, 202-217, 221-228, 259-265, 275-290, 293-303, 309-325, 335-343, 345-351, 365-379, 384-394, 406-414, 423-437, 452-465, 478-507, 525-534, 554-560, 611-624, 628-651, 669-682, 742-747, 767-778, 782-792, 804-812, 820-836, 79-231 and 359-451 of Seq ID No 189; 5-28, 39-45, 56-62, 67-74, 77-99, 110-117, 124-141, 168-176, 200-230, 237-244, 268-279, 287-299, 304-326, 329-335, 348-362, 370-376, 379-384, 390-406, 420-429, 466-471, 479-489, 495-504, 529-541, 545-553, 561-577, 598-604, 622-630, 637-658, 672-680, 682-688, 690-696, 698-709, 712-719, 724-736, 738-746, 759-769, 780-786, 796-804, 813-818, 860-877, 895-904, 981-997, 1000-1014, 1021-1029, 1-162, 206-224, 254-350, 414-514 and 864-938 of Seq ID No 190; 4-11, 19-49, 56-66, 68-101, 109-116, 123-145, 156-165, 177-185, 204-221, 226-234, 242-248, 251-256, 259-265, 282-302, 307-330, 340-349, 355-374, 377-383, 392-400, 422-428, 434-442, 462-474 and 266-322 of Seq ID No 191; 14-43, 45-57, 64-74, 80-87, 106-127, 131-142, 145-161, 173-180, 182-188, 203-210, 213-219, 221-243, 245-254, 304-311, 314-320, 342-348, 354-365, 372-378, 394-399, 407-431, 436-448, 459-465, 470-477, 484-490, 504-509, 531-537, 590-596, 611-617, 642-647, 723-734, 740-751, 754-762, 764-774, 782-797, 807-812, 824-831, 838-845, 877-885, 892-898, 900-906, 924-935, 940-946, 982-996, 1006-1016, 1033-1043, 1051-1056, 1058-1066, 1094-1108, 1119-1126, 1129-1140, 1150-1157, 1167-1174, 1176-1185, 1188-1201, 1209-1216, 1220-1228, 1231-1237, 1243-1248, 1253-1285, 1288-1297, 1299-1307, 1316-1334, 1336-1343, 1350-1359, 1365-1381, 1390-1396, 1412-1420, 1427-1439, 1452-1459, 1477-1484, 1493-1512, 1554-1559, 1570-1578, 1603-1608, 1623-1630, 1654-1659, 1672-1680, 1689-1696, 1705-1711, 1721-1738, 1752-1757, 1773-1780, 1817-1829, 1844-1851, 1856-1863, 1883-1895, 1950-1958, 1974-1990, 172-354, 384-448, 464-644, 648-728 and 1357-1370 of Seq ID No 192; 8-27, 68-74, 77-99, 110-116, 124-141, 169-176, 201-216, 220-227, 258-264, 274-289, 292-302, 308-324, 334-342, 344-350, 364-372, 377-387, 399-407, 416-429, 445-458, 471-481, 483-500, 518-527, 547-553, 604-617, 621-644, 662-675, 767-778, 809-816, 15-307, 350-448 and 496-620 of Seq ID No 193; 4-17, 24-29, 53-59, 62-84, 109-126, 159-164, 189-204, 208-219, 244-249, 274-290, 292-302, 308-324, 334-342, 344-350, 378-389, 391-397, 401-409, 424-432, 447-460, 470-479, 490-504, 521-529, 538-544, 549-555, 570-577, 583-592, 602-608, 615-630, 635-647, 664-677, 692-698, 722-731, 733-751, 782-790, 793-799, 56-267, 337-426 and 495-601 of Seq ID No 194; 12-22, 49-59, 77-89, 111-121, 136-148, 177-186, 207-213, 217-225, 227-253, 259-274, 296-302, 328-333, 343-354, 374-383, 424-446, 448-457, 468-480, 488-502, 507-522, 544-550, 553-560, 564-572, 587-596, 604-614, 619-625, 629-635, 638-656, 662-676, 680-692, 697-713, 720-738, 779-786, 833-847, 861-869, 880-895, 897-902, 911-917, 946-951, 959-967, 984-990, 992-1004, 1021-1040, 1057-1067, 1073-1080 and 381-403 of Seq ID No 195; 4-10, 26-31, 46-56, 60-66, 70-79, 86-94, 96-102, 109-118, 132-152, 164-187, 193-206, 217-224 and 81-149 of Seq ID No 196; 4-21, 26-37, 48-60, 71-82, 109-117, 120-128, 130-136, 142-147, 181-187, 203-211, 216-223, 247-255, 257-284, 316-325, 373-379, 395-400, 423-435, 448-456, 479-489, 512-576, 596-625, 641-678, 680-688, 692-715 and 346-453 of Seq ID No 197; 10-16, 25-31, 34-56, 58-69, 71-89, 94-110, 133-176, 186-193, 208-225, 240-250, 259-266, 302-307, 335-341, 376-383, 410-416 and 316-407 of Seq ID No 198; 11-29, 42-56, 60-75, 82-88, 95-110, 116-126, 132-143, 145-160, 166-172, 184-216 and 123-164 of Seq ID No 199; 11-29, 54-63, 110-117, 139-152, 158-166, 172-180, 186-193, 215-236, 240-251, 302-323, 330-335, 340-347, 350-366, 374-381 and 252-299 of Seq ID No 200; 18-27, 35-42, 50-56, 67-74, 112-136, 141-153, 163-171, 176-189, 205-213, 225-234, 241-247, 253-258, 269-281, 288-298, 306-324, 326-334, 355-369, 380-387 and 289-320 of Seq ID No 201; 7-15, 19-41, 56-72, 91-112, 114-122, 139-147, 163-183, 196-209, 258-280, 326-338, 357-363, 391-403, 406-416 and 360-378 of Seq ID No 202; 11-18, 29-41, 43-49, 95-108, 142-194, 204-212, 216-242, 247-256, 264-273 and 136-149 of Seq ID No 203; 18-24, 33-40, 65-79, 89-102, 113-119, 130-137, 155-161, 173-179, 183-203, 205-219, 223-231, 245-261, 267-274, 296-306, 311-321, 330-341, 344-363, 369-381, 401-408, 415-427, 437-444, 453-464, 472-478, 484-508, 517-524, 526-532, 543-548 and 59-180 of Seq ID No 204; 5-13, 52-65, 67-73, 97-110, 112-119, 134-155 and 45-177 of Seq ID No 205; 6-28, 34-43, 57-67, 75-81, 111-128, 132-147, 155-163, 165-176, 184-194, 208-216, 218-229, 239-252, 271-278, 328-334, 363-376, 381-388, 426-473, 481-488, 492-498, 507-513, 536-546, 564-582, 590-601, 607-623, 148-269, 420-450 and 610-648 of Seq ID No 206; 4-12, 20-38, 69-75, 83-88, 123-128, 145-152, 154-161, 183-188, 200-213, 245-250, 266-272, 306-312, 332-339, 357-369, 383-389, 395-402, 437-453, 455-470, 497-503 and 1-112 of Seq ID No 207; 35-59, 74-86, 111-117, 122-137 and 70-154 of Seq ID No 208; 26-42, 54-61, 65-75, 101-107, 123-130, 137-144, 148-156, 164-172, 177-192, 213-221, 231-258 and 157-249 of Seq ID No 209; 29-38, 61-67, 77-87, 94-100, 105-111, 118-158 and 1-97 of Seq ID No 210; 7-21, 30-48, 51-58, 60-85, 94-123, 134-156, 160-167, 169-183, 186-191, 216-229, 237-251,

257-267, 272-282, 287-298 and 220-243 of Seq ID No 211; 6-29, 34-47, 56-65, 69-76, 83-90, 123-134, 143-151, 158-178, 197-203, 217-235, 243-263, 303-309, 320-333, 338-348, 367-373, 387-393, 407-414, 416-427, 441-457, 473-482, 487-499, 501-509, 514-520, 530-535, 577-583, 590-602, 605-612, 622-629, 641-670, 678-690, 37-71 and 238-307 of Seq ID No 212; 7-40, 121-132, 148-161, 196-202, 209-215, 221-235, 248-255, 271-280, 288-295, 330-339, 395-409, 414-420, 446-451, 475-487, 556-563, 568-575, 580-586, 588-595, 633-638, 643-648, 652-659, 672-685, 695-700, 710-716, 737-742, 749-754, 761-767, 775-781, 796-806, 823-835, 850-863, 884-890, 892-900, 902-915, 934-941 and 406-521 of Seq ID No 213; 9-18, 24-46, 51-58, 67-77, 85-108, 114-126, 129-137, 139-146, 152-165, 173-182, 188-195, 197-204, 217-250, 260-274, 296-313, 343-366, 368-384, 427-434, 437-446, 449-455, 478-484, 492-506, 522-527, 562-591, 599-606, 609-618, 625-631, 645-652 and 577-654 of Seq ID No 214; 13-20, 26-37, 41-53, 56-65, 81-100, 102-114, 118-127, 163-188, 196-202, 231-238, 245-252, 266-285, 293-298, 301-306 and 19-78 of Seq ID No 215; 10-23, 32-42, 54-66, 73-91, 106-113, 118-127, 139-152, 164-173, 198-207, 210-245, 284-300, 313-318, 330-337, 339-346, 354-361, 387-393, 404-426, 429-439, 441-453, 467-473, 479-485, 496-509, 536-544, 551-558, 560-566, 569-574, 578-588, 610-615, 627-635, 649-675, 679-690, 698-716, 722-734, 743-754, 769-780, 782-787 and 480-550 of Seq ID No 216; 6-39, 42-50, 60-68, 76-83, 114-129, 147-162, 170-189, 197-205, 217-231, 239-248, 299-305, 338-344, 352-357, 371-377, 380-451, 459-483, 491-499, 507-523, 537-559, 587-613, 625-681, 689-729, 737-781, 785-809, 817-865, 873-881, 889-939, 951-975, 983-1027, 1031-1055, 1063-1071, 1079-1099, 1103-1127, 1151-1185, 1197-1261, 1269-1309, 1317-1333, 1341-1349, 1357-1465, 1469-1513, 1517-1553, 1557-1629, 1637-1669, 1677-1701, 1709-1725, 1733-1795, 1823-1849, 1861-1925, 1933-1973, 1981-2025, 2029-2053, 2061-2109, 2117-2125, 2133-2183, 2195-2219, 2227-2271, 2275-2299, 2307-2315, 2323-2343, 2347-2371, 2395-2429, 2441-2529, 2537-2569, 2577-2601, 2609-2625, 2633-2695, 2699-2737, 2765-2791, 2803-2867, 2889-2913, 2921-2937, 2945-2969, 2977-2985, 2993-3009, 3023-3045, 3073-3099, 3111-3167, 3175-3215, 3223-3267, 3271-3295, 3303-3351, 3359-3367, 3375-3425, 3437-3461, 3469-3513, 3517-3541, 3549-3557, 3565-3585, 3589-3613, 3637-3671, 3683-3747, 3755-3795, 3803-3819, 3827-3835, 3843-3951, 3955-3999, 4003-4039, 4043-4115, 4123-4143, 4147-4171, 4195-4229, 4241-4305, 4313-4353, 4361-4377, 4385-4393, 4401-4509, 4513-4557, 4561-4597, 4601-4718, 4749-4768, 74-171, 452-559 and 2951-3061 of Seq ID No 217; 16-22, 30-51, 70-111, 117-130, 137-150, 171-178, 180-188, 191-196 and 148-181 of Seq ID No 218; 6-19, 21-46, 50-56, 80-86, 118-126, 167-186, 189-205, 211-242, 244-267, 273-286, 290-297, 307-316, 320-341 and 34-60 of Seq ID No 219; 5-26, 33-43, 48-54, 58-63, 78-83, 113-120, 122-128, 143-152, 157-175, 185-192, 211-225, 227-234, 244-256, 270-281, 284-290, 304-310, 330-337, 348-355, 362-379, 384-394, 429-445, 450-474, 483-490, 511-520, 537-546, 548-554, 561-586, 590-604, 613-629, 149-186, 285-431 and 573-659 of Seq ID No 220; 5-26, 49-59, 61-67, 83-91, 102-111, 145-157, 185-192, 267-272, 279-286, 292-298, 306-312, 134-220, 235-251 and 254-280 of Seq ID No 221; 5-19, 72-79, 83-92, 119-124, 140-145, 160-165, 167-182, 224-232, 240-252, 259-270, 301-310, 313-322, 332-343, 347-367, 384-398, 416-429, 431-446, 454-461 and 1-169 of Seq ID No 222; 8-17, 26-31, 56-62, 75-83, 93-103, 125-131, 135-141, 150-194, 205-217, 233-258, 262-268, 281-286 and 127-168 of Seq ID No 223; 6-12, 69-75, 108-115, 139-159, 176-182, 194-214 and 46-161 of Seq ID No 224; 6-13, 18-27, 39-48, 51-59, 66-73, 79-85, 95-101, 109-116, 118-124, 144-164, 166-177, 183-193, 197-204, 215-223, 227-236, 242-249, 252-259, 261-270, 289-301, 318-325 and 12-58 of Seq ID No 225; 4-10, 26-32, 48-60, 97-105, 117-132, 138-163, 169-185, 192-214, 219-231, 249-261, 264-270, 292-308, 343-356, 385-392, 398-404, 408-417, 435-441 and 24-50 of Seq ID No 226; 10-40, 42-48, 51-61, 119-126 and 1-118 of Seq ID No 227; 5-17, 40-58, 71-83, 103-111, 123-140, 167-177, 188-204 and 116-128 of Seq ID No 228; 4-9, 11-50, 57-70, 112-123, 127-138 and 64-107 of Seq ID No 229; 9-39, 51-67 and 1-101 of Seq ID No 230; 5-14, 17-25, 28-46, 52-59, 85-93, 99-104, 111-120, 122-131, 140-148, 158-179, 187-197, 204-225, 271-283, 285-293 and 139-155 of Seq ID No 231; 42-70, 73-90, 92-108, 112-127, 152-164, 166-172, 181-199, 201-210, 219-228, 247-274, 295-302, 322-334, 336-346, 353-358, 396-414, 419-425, 432-438, 462-471, 518-523, 531-536, 561-567, 576-589, 594-612, 620-631, 665-671, 697-710, 718-731, 736-756, 765-771, 784-801 and 626-653 of Seq ID No 232; 8-28, 41-51, 53-62, 68-74, 79-85, 94-100, 102-108, 114-120, 130-154, 156-162, 175-180, 198-204, 206-213, 281-294, 308-318, 321-339, 362-368, 381-386, 393-399, 407-415 and 2-13 of Seq ID No 233; 4-39, 48-65, 93-98, 106-112, 116-129 and 10-36 of Seq ID No 234; 25-32, 35-50, 66-71, 75-86, 90-96, 123-136, 141-151, 160-179, 190-196, 209-215, 222-228, 235-242, 257-263, 270-280 and 209-247 of Seq ID No 235; 5-29, 31-38, 50-57, 62-75, 83-110, 115-132, 168-195, 197-206, 216-242, 249-258, 262-269, 333-340, 342-350, 363-368, 376-392, 400-406, 410-421, 423-430, 436-442, 448-454, 460-466, 471-476, 491-496, 511-516, 531-536, 551-556, 571-576, 585-591, 599-605, 27-70, 219-293, 441-504 and 512-584 of Seq ID No 236; 4-12, 14-34, 47-75, 83-104, 107-115, 133-140, 148-185, 187-196, 207-212, 224-256, 258-265, 281-287, 289-296,

298-308, 325-333, 345-355, 365-371, 382-395, 424-435, 441-457, 465-472, 483-491, 493-505, 528-534, 536-546, 552-558, 575-584, 589-600, 616-623 and 576-591 of Seq ID No 237; 4-76, 78-89, 91-126, 142-148, 151-191, 195-208, 211-223, 226-240, 256-277, 279-285, 290-314, 317-323, 358-377, 381-387, 391-396, 398-411, 415-434, 436-446, 454-484, 494-512, 516-523, 538-552, 559-566, 571-577, 579-596, 599-615, 620-627, 635-644, 694-707, 720-734, 737-759, 761-771 and 313-329 of Seq ID No 238; 7-38, 44-49, 79-89, 99-108, 117-123, 125-132, 137-146, 178-187, 207-237, 245-255, 322-337, 365-387, 398-408, 445-462, 603-608, 623-628, 644-650, 657-671, 673-679 and 111-566 of Seq ID No 239; 6-20, 22-35, 39-45, 58-64, 77-117, 137-144, 158-163, 205-210, 218-224, 229-236, 239-251, 263-277, 299-307, 323-334, 353-384, 388-396, 399-438, 443-448, 458-463, 467-478, 481-495, 503-509, 511-526, 559-576, 595-600, 612-645, 711-721, 723-738, 744-758, 778-807 and 686-720 of Seq ID No 240; 10-33, 35-41, 72-84, 129-138, 158-163, 203-226, 243-252, 258-264, 279-302, 322-329, 381-386, 401-406, 414-435 and 184-385 of Seq ID No 241; 4-9, 19-24, 41-47, 75-85, 105-110, 113-146 and 45-62 of Seq ID No 242; 4-25, 52-67, 117-124, 131-146, 173-180, 182-191, 195-206, 215-221, 229-236, 245-252, 258-279, 286-291, 293-302, 314-320, 327-336, 341-353, 355-361, 383-389 and 1-285 of Seq ID No 243; 14-32, 38-50, 73-84, 93-105, 109-114 and 40-70 of Seq ID No 244; 5-26 and 22-34 of Seq ID No 245; 23-28 and 13-39 of Seq ID No 246; 8-14 and 21-34 of Seq ID No 247; 4-13, 20-29, 44-50, 59-74 and 41-69 of Seq ID No 248; 4-9, 19-42, 48-59, 71-83 and 57-91 of Seq ID No 249; 4-14 and 10-28 of Seq ID No 250; 22-28, 32-42, 63-71, 81-111, 149-156, 158-167, 172-180, 182-203, 219-229 and 27-49 of Seq ID No 251; 17-27 and 23-32 of Seq ID No 252; 18-24 and 28-38 of Seq ID No 253; 9-15 and 13-27 of Seq ID No 254; 13-22 and 18-29 of Seq ID No 255; 17-26 and 2-11 of Seq ID No 256; 4-33 and 16-32 of Seq ID No 257; 4-10, 37-43, 54-84, 92-127 and 15-62 of Seq ID No 258; 4-14, 20-32, 35-60, 69-75, 79-99, 101-109, 116-140 and 124-136 of Seq ID No 259; 2-13 of Seq ID No 260; 4-13, 28-42 and 42-57 of Seq ID No 261; 4-14, 27-44 and 14-35 of Seq ID No 262; 4-12 and 1-27 of Seq ID No 263; 4-18, 39-45, 47-74 and 35-66 of Seq ID No 264; 8-20, 43-77 and 17-36 of Seq ID No 265; 4-30, 35-45, 51-57 and 35-49 of Seq ID No 266; 4-24, 49-57 and 15-34 of Seq ID No 267; 4-22 and 8-27 of Seq ID No 268; 13-25, 32-59, 66-80 and 21-55 of Seq ID No 269; 4-10, 24-33, 35-42, 54-65, 72-82, 98-108 and 15-30 of Seq ID No 270; 8-19 and 17-47 of Seq ID No 271; 12-18, 40-46 and 31-52 of Seq ID No 272; 4-20, 35-78, 83-102, 109-122 and 74-86 of Seq ID No 273; 7-17, 21-41, 46-63 and 2-20 of Seq ID No 274; 30-37 and 2-33 of Seq ID No 275; 4-13, 17-25 and 1-15 of Seq ID No 276; 17-31, 44-51 and 20-51 of Seq ID No 277; 20-30 and 5-23 of Seq ID No 278; 13-33, 48-71 and 92-110 of Seq ID No 279; 4-9, 50-69, 76-88, 96-106, 113-118 and 12-34 of Seq ID No 280; 4-24 and 6-26 of Seq ID No 281; 7-26 and 14-30 of Seq ID No 282; 9-39, 46-68, 75-82, 84-103 and 26-44 of Seq ID No 283; 4-30, 33-107 and 58-84 of Seq ID No 284; 4-12 and 9-51 of Seq ID No 285; 12-18, 29-37 and 6-37 of Seq ID No 286; 4-21, 33-52, 64-71 and 16-37 of Seq ID No 287; 9-19 and 2-30 of Seq ID No 288; 20-37 of Seq ID No 245; 8-27 of Seq ID No 246; 10-27 of Seq ID No 247; 42-59 and 52-69 of Seq ID No 248; 63-80 and 74-91 of Seq ID No 249; 11-28 of Seq ID No 250; 28-49 of Seq ID No 251; 15-32 of Seq ID No 252; 4-20 of Seq ID No 253; 10-27 of Seq ID No 254; 17-34 of Seq ID No 255; 1-18 of Seq ID No 256; 16-33 of Seq ID No 257; 16-36, 30-49 and 43-62 of Seq ID No 258; 122-139 of Seq ID No 259; 1-18 of Seq ID No 260; 41-58 of Seq ID No 261; 15-35 of Seq ID No 262; 2-27 of Seq ID No 263; 18-36 of Seq ID No 265; 34-51 of Seq ID No 266; 9-27 of Seq ID No 268; 22-47 of Seq ID No 269; 18-36 and 29-47 of Seq ID No 271; 32-52 of Seq ID No 272; 72-89 of Seq ID No 273; 3-20 of Seq ID No 274; 3-21 and 15-33 of Seq ID No 275; 1-18 of Seq ID No 276; 6-23 of Seq ID No 278; 93-110 of Seq ID No 279; 13-34 of Seq ID No 280; 7-26 and 9-26 of Seq ID No 281; 16-33 of Seq ID No 282; 27-44 of Seq ID No 283; 67-84 of Seq ID No 284; 10-33 and 26-50 of Seq ID No 285; 7-25 and 19-37 of Seq ID No 286; 17-37 of Seq ID No 287; 3-20 and 13-30 of Seq ID No 288; 62-80 and 75-93 of Seq ID No 145; 92-108 of Seq ID No 147; 332-349, 177-200 and 1755-1777 of Seq ID No 148; 109-133, 149-174, 260-285 and 460-485 of Seq ID No 149; 26-47 and 42-64 of Seq ID No 150; 22-41, 35-54, 115-130, 306-325, 401-420 and 454-478 of Seq ID No 151; 22-45 of Seq ID No 155; 156-174, 924-940, 1485-1496, 1447-1462 and 1483-1498 of Seq ID No 160; 457-475 of Seq ID No 161; 302-325 of Seq ID No 163; 288-305 of Seq ID No 164; 244-266 and 260-282 of Seq ID No 165; 204-225 and 220-241 of Seq ID No 166; 324-345, 340-361, 356-377, 372-393 and 388-408 of Seq ID No 167; 39-64 of Seq ID No 168; 54-76 and 70-92 of Seq ID No 169; 1227-1247, 1539-1559, 1554-1574, 1569-1589, 1584-1604, 1242-1262, 1272-1292, 1287-1308, 1456-1477, 1472-1494, 1488-1510 and 1505-1526 of Seq ID No 170; 351-368 of Seq ID No 172; 179-200, 195-216, 211-232, 227-248 and 243-263 of Seq ID No 173; 13-37 of Seq ID No 174; 208-224 of Seq ID No 175; 42-64, 59-81, 304-328, 323-348, 465-489, 968-992, 1399-1418, 1412-1431 and 2092-2111 of Seq ID No 177; 1528-1547 and 1541-1560 of Seq ID No

178; 184-200, 367-388, 382-403, 409-429, 425-444 and 438-457 of Seq ID No 179; 27-50 and 45-67 of Seq ID No 180; 114-131 and 405-419 of Seq ID No 183; 113-134, 129-150, 145-166, 161-182 and 177-198 of Seq ID No 184; 495-515 of Seq ID No 186; 346-358 of Seq ID No 187; 208-224 of Seq ID No 190; 178-194, 202-223, 217- 238, 288-308 and 1355-1372 of Seq ID No 192; 57-78 of Seq ID No 194; 347-369, 364-386, 381-403, 398-420, 415-437 and 432-452 of Seq ID No 197; 347-372 of Seq ID No 198; 147-163 of Seq ID No 199; 263-288 of Seq ID No 200; 361-377 of Seq ID No 202; 82-104, 99-121, 116-138, 133-155 and 150-171 of Seq ID No 204; 110-130 and 125-145 of Seq ID No 205; 613-631, 626-644 and 196-213 of Seq ID No 206; 78-100, 95-117, 112-134 and 129-151 of Seq ID No 208; 158-180, 175-197, 192-214, 209-231 and 226-248 of Seq ID No 209; 30-50, 45-65 and 60-79 of Seq ID No 210; 431-455 and 450-474 of Seq ID No 213; 579-601, 596-618, 613-635 and 630-653 of Seq ID No 214; 920-927, 98-119, 114-135, 130-151, 146-167 and 162-182 of Seq ID No 217; 36-59 of Seq ID No 219; 194-216 and 381-404 of Seq ID No 220; 236-251 and 255-279 of Seq ID No 221; 80-100 and 141-164 of Seq ID No 222; 128-154 of Seq ID No 223; 82-100, 95-116 and 111-134 of Seq ID No 224; 55-76, 71-92 and 87-110 of Seq ID No 227; 91-106 of Seq ID No 229; 74-96 of Seq ID No 230; 140-157 of Seq ID No 231; 4-13 of Seq ID No 233; 41-65 and 499-523 of Seq ID No 236; 122-146, 191-215, 288-313, 445-469 and 511-535 of Seq ID No 239; 347-368 of Seq ID No 241; 46-61 of Seq ID No 242; 15-37, 32-57, 101-121, 115-135, 138-158, 152-172, 220-242 and 236-258 of Seq ID No 243; and fragments comprising at least 6, preferably more than 8, especially more than 10 aa and preferably not more than 70, 50, 40, 20, 15, 11 aa of said sequences. All these fragments individually and each independently form a preferred selected aspect of the present invention.

All linear hyperimmune serum reactive fragments of a particular antigen may be identified by analysing the entire sequence of the protein antigen by a set of peptides overlapping by 1 amino acid with a length of at least 10 amino acids. Subsequently, non-linear epitopes can be identified by analysis of the protein antigen with hyperimmune sera using the expressed full-length protein or domain polypeptides thereof. Assuming that a distinct domain of a protein is sufficient to form the 3D structure independent from the native protein, the analysis of the respective recombinant or synthetically produced domain polypeptide with hyperimmune serum would allow the identification of conformational epitopes within the individual domains of multi-domain proteins. For those antigens where a domain possesses linear as well as conformational epitopes, competition experiments with peptides corresponding to the linear epitopes may be used to confirm the presence of conformational epitopes.

It will be appreciated that the invention also relates to, among others, nucleic acid molecules encoding the aforementioned fragments, nucleic acid molecules that hybridise to nucleic acid molecules encoding the fragments, particularly those that hybridise under stringent conditions, and nucleic acid molecules, such as PCR primers, for amplifying nucleic acid molecules that encode the fragments. In these regards, preferred nucleic acid molecules are those that correspond to the preferred fragments, as discussed above.

The present invention also relates to vectors, which comprise a nucleic acid molecule or nucleic acid molecules of the present invention, host cells which are genetically engineered with vectors of the invention and the production of hyperimmune serum reactive antigens and fragments thereof by recombinant techniques.

A great variety of expression vectors can be used to express a hyperimmune serum reactive antigen or fragment thereof according to the present invention. Generally, any vector suitable to maintain, propagate or express nucleic acids to express a polypeptide in a host may be used for expression in this regard. In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Starting plasmids disclosed herein are either commercially available, publicly available, or can be constructed from available plasmids by routine application of well-known, published procedures. Preferred among vectors, in certain respects, are those for expression of nucleic acid molecules and hyperimmune serum reactive antigens or fragments thereof of the present invention. Nucleic acid constructs in host cells can

be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the hyperimmune serum reactive antigens and fragments thereof of the invention can be synthetically produced by conventional peptide synthesizers. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA construct of the present invention.

Host cells can be genetically engineered to incorporate nucleic acid molecules and express nucleic acid molecules of the present invention. Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli, Streptomyces* and *Bacillus subtillis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, Hela, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

The invention also provides a process for producing a *S. pneumoniae* hyperimmune serum reactive antigen and a fragment thereof comprising expressing from the host cell a hyperimmune serum reactive antigen or fragment thereof encoded by the nucleic acid molecules provided by the present invention. The invention further provides a process for producing a cell, which expresses a *S. pneumoniae* hyperimmune serum reactive antigen or a fragment thereof comprising transforming or transfecting a suitable host cell with the vector according to the present invention such that the transformed or transfected cell expresses the polypeptide encoded by the nucleic acid contained in the vector.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N- or C-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, regions may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability or to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize or purify polypeptides. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobin molecules together with another protein or part thereof. In drug discovery, for example, proteins have been fused with antibody Fc portions for the purpose of high-throughout screening assays to identify antagonists. See for example, {Bennett, D. et al., 1995} and {Johanson, K. et al., 1995}.

The *S. pneumoniae* hyperimmune serum reactive antigen or a fragment thereof can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, hydroxylapatite chromatography and lectin chromatography.

The hyperimmune serum reactive antigens and fragments thereof according to the present invention can be produced by chemical synthesis as well as by biotechnological means. The latter comprise the transfection or transformation of a host cell with a vector containing a nucleic acid according to the present invention and the cultivation of the transfected or transformed host cell under conditions, which are known to the ones skilled in the art. The production method may also comprise a purification step in order to purify or isolate the polypeptide to be manufactured. In a preferred embodiment the vector is a vector according to the present invention.

The hyperimmune serum reactive antigens and fragments thereof according to the present invention may be used for the detection of the organism or organisms in a sample containing these organisms or

polypeptides derived thereof. Preferably such detection is for diagnosis, more preferable for the diagnosis of a disease, most preferably for the diagnosis of a diseases related or linked to the presence or abundance of Gram-positive bacteria, especially bacteria selected from the group comprising streptococci, staphylococci and lactococci. More preferably, the microorganisms are selected from the group comprising *Streptococcus agalactiae, Streptococcus pyogenes* and *Streptococcus mutans*, especially the microorganism is *Streptococcus pyogenes*.

The present invention also relates to diagnostic assays such as quantitative and diagnostic assays for detecting levels of the hyperimmune serum reactive antigens and fragments thereof of the present invention in cells and tissues, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for detecting over-expression of the polypeptide compared to normal control tissue samples may be used to detect the presence of an infection, for example, and to identify the infecting organism. Assay techniques that can be used to determine levels of a polypeptide, in a sample derived from a host are well known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these, ELISAs frequently are preferred. An ELISA assay initially comprises preparing an antibody specific to the polypeptide, preferably a monoclonal antibody. In addition, a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached to a detectable reagent such as radioactive, fluorescent or enzymatic reagent, such as horseradish peroxidase enzyme.

The hyperimmune serum reactive antigens and fragments thereof according to the present invention may also be used for the purpose of or in connection with an array. More particularly, at least one of the hyperimmune serum reactive antigens and fragments thereof according to the present invention may be immobilized on a support. Said support typically comprises a variety of hyperimmune serum reactive antigens and fragments thereof whereby the variety may be created by using one or several of the hyperimmune serum reactive antigens and fragments thereof according to the present invention and/or hyperimmune serum reactive antigens and fragments thereof being different. The characterizing feature of such array as well as of any array in general is the fact that at a distinct or predefined region or position on said support or a surface thereof, a distinct polypeptide is immobilized. Because of this any activity at a distinct position or region of an array can be correlated with a specific polypeptide. The number of different hyperimmune serum reactive antigens and fragments thereof immobilized on a support may range from as little as 10 to several 1000 different hyperimmune serum reactive antigens and fragments thereof. The density of hyperimmune serum reactive antigens and fragments thereof per cm² is in a preferred embodiment as little as 10 peptides/polypeptides per cm² to at least 400 different peptides/polypeptides per cm² and more particularly at least 1000 different hyperimmune serum reactive antigens and fragments thereof per cm².

The manufacture of such arrays is known to the one skilled in the art and, for example, described in US patent 5,744,309. The array preferably comprises a planar, porous or non-porous solid support having at least a first surface. The hyperimmune serum reactive antigens and fragments thereof as disclosed herein, are immobilized on said surface. Preferred support materials are, among others, glass or cellulose. It is also within the present invention that the array is used for any of the diagnostic applications described herein. Apart from the hyperimmune serum reactive antigens and fragments thereof according to the present invention also the nucleic acid molecules according to the present invention may be used for the generation of an array as described above. This applies as well to an array made of antibodies, preferably monoclonal antibodies as, among others, described herein.

In a further aspect the present invention relates to an antibody directed to any of the hyperimmune serum reactive antigens and fragments thereof, derivatives or fragments thereof according to the present invention. The present invention includes, for example, monoclonal and polyclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of a Fab expression

library. It is within the present invention that the antibody may be chimeric, i. e. that different parts thereof stem from different species or at least the respective sequences are taken from different species.

Antibodies generated against the hyperimmune serum reactive antigens and fragments thereof corresponding to a sequence of the present invention can be obtained by direct injection of the hyperimmune serum reactive antigens and fragments thereof into an animal or by administering the hyperimmune serum reactive antigens and fragments thereof to an animal, preferably a non-human. The antibody so obtained will then bind the hyperimmune serum reactive antigens and fragments thereof itself. In this manner, even a sequence encoding only a fragment of a hyperimmune serum reactive antigen and fragments thereof can be used to generate antibodies binding the whole native hyperimmune serum reactive antigen and fragments thereof. Such antibodies can then be used to isolate the hyperimmune serum reactive antigens and fragments thereof from tissue expressing those hyperimmune serum reactive antigens and fragments thereof.

For preparation of monoclonal antibodies, any technique known in the art, which provides antibodies produced by continuous cell line cultures can be used (as described originally in {Kohler, G. et al., 1975}.

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic hyperimmune serum reactive antigens and fragments thereof according to this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies to immunogenic hyperimmune serum reactive antigens and fragments thereof according to this invention.

Alternatively, phage display technology or ribosomal display could be utilized to select antibody genes with binding activities towards the hyperimmune serum reactive antigens and fragments thereof either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing respective target antigens or from naïve libraries {McCafferty, J. et al., 1990}; {Marks, J. et al., 1992}. The affinity of these antibodies can also be improved by chain shuffling {Clackson, T. et al., 1991}.

If two antigen binding domains are present, each domain may be directed against a different epitope – termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the hyperimmune serum reactive antigens and fragments thereof or purify the hyperimmune serum reactive antigens and fragments thereof of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography.

Thus, among others, antibodies against the hyperimmune serum reactive antigens and fragments thereof of the present invention may be employed to inhibit and/or treat infections, particularly bacterial infections and especially infections arising from S. pneumoniae.

Hyperimmune serum reactive antigens and fragments thereof include antigenically, epitopically or immunologically equivalent derivatives, which form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a hyperimmune serum reactive antigen and fragments thereof or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or hyperimmune serum reactive antigen and fragments thereof according to the present invention, interfere with the interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the interaction between pathogen and mammalian host.

The hyperimmune serum reactive antigens and fragments thereof, such as an antigenically or

- 34 -

immunologically equivalent derivative or a fusion protein thereof can be used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the hyperimmune serum reactive antigens and fragments thereof. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein, for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively, an antigenic peptide comprising multiple copies of the protein or hyperimmune serum reactive antigen and fragments thereof, or an antigenically or immunologically equivalent hyperimmune serum reactive antigen and fragments thereof, may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably the antibody or derivative thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized", wherein the complimentarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in {Jones, P. et al., 1986} or {Tempest, P. et al., 1991}.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscle, delivery of DNA complexed with specific protein carriers, coprecipitation of DNA with calcium phosphate, encapsulation of DNA in various forms of liposomes, particle bombardment {Tang, D. et al., 1992}; {Eisenbraun, M. et al., 1993} and *in vivo* infection using cloned retroviral vectors {Seeger, C. et al., 1984}.

In a further aspect the present invention relates to a peptide binding to any of the hyperimmune serum reactive antigens and fragments thereof according to the present invention, and a method for the manufacture of such peptides whereby the method is characterized by the use of the hyperimmune serum reactive antigens and fragments thereof according to the present invention and the basic steps are known to the one skilled in the art.

Such peptides may be generated by using methods according to the state of the art such as phage display or ribosome display. In case of phage display, basically a library of peptides is generated, in form of phages, and this kind of library is contacted with the target molecule, in the present case a hyperimmune serum reactive antigen and fragments thereof according to the present invention. Those peptides binding to the target molecule are subsequently removed, preferably as a complex with the target molecule, from the respective reaction. It is known to the one skilled in the art that the binding characteristics, at least to a certain extent, depend on the particularly realized experimental set-up such as the salt concentration and the like. After separating those peptides binding to the target molecule with a higher affinity or a bigger force, from the non-binding members of the library, and optionally also after removal of the target molecule from the complex of target molecule and peptide, the respective peptide(s) may subsequently be characterised. Prior to the characterisation optionally an amplification step is realized such as, e. g. by propagating the peptide encoding phages. The characterisation preferably comprises the sequencing of the target binding peptides. Basically, the peptides are not limited in their lengths, however, preferably peptides having a lengths from about 8 to 20 amino acids are preferably obtained in the respective methods. The size of the libraries may be about 10^2 to 10^{18}, preferably 10^8 to 10^{15} different peptides, however, is not limited thereto.

A particular form of target binding hyperimmune serum reactive antigens and fragments thereof are the so-called "anticalines" which are, among others, described in German patent application DE 197 42 706.

In a further aspect the present invention relates to functional nucleic acids interacting with any of the hyperimmune serum reactive antigens and fragments thereof according to the present invention, and a method for the manufacture of such functional nucleic acids whereby the method is characterized by the use of the hyperimmune serum reactive antigens and fragments thereof according to the present invention and the basic steps are known to the one skilled in the art. The functional nucleic acids are

preferably aptamers and spiegelmers.

Aptamers are D-nucleic acids, which are either single stranded or double stranded and which specifically interact with a target molecule. The manufacture or selection of aptamers is, e.g. described in European patent EP 0 533 838. Basically the following steps are realized. First, a mixture of nucleic acids, i. e. potential aptamers, is provided whereby each nucleic acid typically comprises a segment of several, preferably at least eight subsequent randomised nucleotides. This mixture is subsequently contacted with the target molecule whereby the nucleic acid(s) bind to the target molecule, such as based on an increased affinity towards the target or with a bigger force thereto, compared to the candidate mixture. The binding nucleic acid(s) are/is subsequently separated from the remainder of the mixture. Optionally, the thus obtained nucleic acid(s) is amplified using, e.g. polymerase chain reaction. These steps may be repeated several times giving at the end a mixture having an increased ratio of nucleic acids specifically binding to the target from which the final binding nucleic acid is then optionally selected. These specifically binding nucleic acid(s) are referred to as aptamers. It is obvious that at any stage of the method for the generation or identification of the aptamers samples of the mixture of individual nucleic acids may be taken to determine the sequence thereof using standard techniques. It is within the present invention that the aptamers may be stabilized such as, e. g., by introducing defined chemical groups which are known to the one skilled in the art of generating aptamers. Such modification may for example reside in the introduction of an amino group at the 2'-position of the sugar moiety of the nucleotides. Aptamers are currently used as therapeutical agents. However, it is also within the present invention that the thus selected or generated aptamers may be used for target validation and/or as lead substance for the development of medicaments, preferably of medicaments based on small molecules. This is actually done by a competition assay whereby the specific interaction between the target molecule and the aptamer is inhibited by a candidate drug whereby upon replacement of the aptamer from the complex of target and aptamer it may be assumed that the respective drug candidate allows a specific inhibition of the interaction between target and aptamer, and if the interaction is specific, said candidate drug will, at least in principle, be suitable to block the target and thus decrease its biological availability or activity in a respective system comprising such target. The thus obtained small molecule may then be subject to further derivatisation and modification to optimise its physical, chemical, biological and/or medical characteristics such as toxicity, specificity, biodegradability and bioavailability.

Spiegelmers and their generation or manufacture is based on a similar principle. The manufacture of spiegelmers is described in international patent application WO 98/08856. Spiegelmers are L-nucleic acids, which means that they are composed of L-nucleotides rather than D-nucleotides as aptamers are. Spiegelmers are characterized by the fact that they have a very high stability in biological systems and, comparable to aptamers, specifically interact with the target molecule against which they are directed. In the process of generating spiegelmers, a heterogeonous population of D-nucleic acids is created and this population is contacted with the optical antipode of the target molecule, in the present case for example with the D-enantiomer of the naturally occurring L-enantiomer of the hyperimmune serum reactive antigens and fragments thereof according to the present invention. Subsequently, those D-nucleic acids are separated which do not interact with the optical antipode of the target molecule. But those D-nucleic acids interacting with the optical antipode of the target molecule are separated, optionally identified and/or sequenced and subsequently the corresponding L-nucleic acids are synthesized based on the nucleic acid sequence information obtained from the D-nucleic acids. These L-nucleic acids which are identical in terms of sequence with the aforementioned D-nucleic acids interacting with the optical antipode of the target molecule, will specifically interact with the naturally occurring target molecule rather than with the optical antipode thereof. Similar to the method for the generation of aptamers it is also possible to repeat the various steps several times and thus to enrich those nucleic acids specifically interacting with the optical antipode of the target molecule.

In a further aspect the present invention relates to functional nucleic acids interacting with any of the nucleic acid molecules according to the present invention, and a method for the manufacture of such

functional nucleic acids whereby the method is characterized by the use of the nucleic acid molecules and their respective sequences according to the present invention and the basic steps are known to the one skilled in the art. The functional nucleic acids are preferably ribozymes, antisense oligonucleotides and siRNA.

Ribozymes are catalytically active nucleic acids which preferably consist of RNA which basically comprises two moieties. The first moiety shows a catalytic activity whereas the second moiety is responsible for the specific interaction with the target nucleic acid, in the present case the nucleic acid coding for the hyperimmune serum reactive antigens and fragments thereof according to the present invention. Upon interaction between the target nucleic acid and the second moiety of the ribozyme, typically by hybridisation and Watson-Crick base pairing of essentially complementary stretches of bases on the two hybridising strands, the catalytically active moiety may become active which means that it catalyses, either intramolecularly or intermolecularly, the target nucleic acid in case the catalytic activity of the ribozyme is a phosphodiesterase activity. Subsequently, there may be a further degradation of the target nucleic acid, which in the end results in the degradation of the target nucleic acid as well as the protein derived from the said target nucleic acid. Ribozymes, their use and design principles are known to the one skilled in the art, and, for example described in {Doherty, E. et al., 2001} and {Lewin, A. et al., 2001}.

The activity and design of antisense oligonucleotides for the manufacture of a medicament and as a diagnostic agent, respectively, is based on a similar mode of action. Basically, antisense oligonucleotides hybridise based on base complementarity, with a target RNA, preferably with a mRNA, thereby activating RNase H. RNase H is activated by both phosphodiester and phosphorothioate-coupled DNA. Phosphodiester-coupled DNA, however, is rapidly degraded by cellular nucleases with the exception of phosphorothioate-coupled DNA. These resistant, non-naturally occurring DNA derivatives do not inhibit RNase H upon hybridisation with RNA. In other words, antisense polynucleotides are only effective as DNA RNA hybride complexes. Examples for this kind of antisense oligonucleotides are described, among others, in US-patent US 5,849,902 and US 5,989,912. In other words, based on the nucleic acid sequence of the target molecule which in the present case are the nucleic acid molecules for the hyperimmune serum reactive antigens and fragments thereof according to the present invention, either from the target protein from which a respective nucleic acid sequence may in principle be deduced, or by knowing the nucleic acid sequence as such, particularly the mRNA, suitable antisense oligonucleotides may be designed base on the principle of base complementarity.

Particularly preferred are antisense-oligonucleotides, which have a short stretch of phosphorothioate DNA (3 to 9 bases). A minimum of 3 DNA bases is required for activation of bacterial RNase H and a minimum of 5 bases is required for mammalian RNase H activation. In these chimeric oligonucleotides there is a central region that forms a substrate for RNase H that is flanked by hybridising "arms" comprised of modified nucleotides that do not form substrates for RNase H. The hybridising arms of the chimeric oligonucleotides may be modified such as by 2'-O-methyl or 2'-fluoro. Alternative approaches used methylphosphonate or phosphoramidate linkages in said arms. Further embodiments of the antisense oligonucleotide useful in the practice of the present invention are P-methoxyoligonucleotides, partial P-methoxyoligodeoxyribonucleotides or P-methoxyoligonucleotides.

Of particular relevance and usefulness for the present invention are those antisense oligonucleotides as more particularly described in the above two mentioned US patents. These oligonucleotides contain no naturally occurring 5'→3'-linked nucleotides. Rather the oligonucleotides have two types of nucleotides: 2'-deoxyphosphorothioate, which activate RNase H, and 2'-modified nucleotides, which do not. The linkages between the 2'-modified nucleotides can be phosphodiesters, phosphorothioate or P-ethoxyphosphodiester. Activation of RNase H is accomplished by a contiguous RNase H-activating region, which contains between 3 and 5 2'-deoxyphosphorothioate nucleotides to activate bacterial RNase H and between 5 and 10 2'- deoxyphosphorothioate nucleotides to activate eucaryotic and, particularly,

mammalian RNase H. Protection from degradation is accomplished by making the 5′ and 3′ terminal bases highly nuclease resistant and, optionally, by placing a 3′ terminal blocking group.

More particularly, the antisense oligonucleotide comprises a 5′ terminus and a 3′ terminus; and from position 11 to 59 5′→3′-linked nucleotides independently selected from the group consisting of 2′-modified phosphodiester nucleotides and 2′-modified P-alkyloxyphosphotriester nucleotides; and wherein the 5′-terminal nucleoside is attached to an RNase H-activating region of between three and ten contiguous phosphorothioate-linked deoxyribonucleotides, and wherein the 3′-terminus of said oligonucleotide is selected from the group consisting of an inverted deoxyribonucleotide, a contiguous stretch of one to three phosphorothioate 2′-modified ribonucleotides, a biotin group and a P-alkyloxyphosphotriester nucleotide.

Also an antisense oligonucleotide may be used wherein not the 5′ terminal nucleoside is attached to an RNase H-activating region but the 3′ terminal nucleoside as specified above. Also, the 5′ terminus is selected from the particular group rather than the 3′ terminus of said oligonucleotide.

The nucleic acids as well as the hyperimmune serum reactive antigens and fragments thereof according to the present invention may be used as or for the manufacture of pharmaceutical compositions, especially vaccines. Preferably such pharmaceutical composition, preferably vaccine is for the prevention or treatment of diseases caused by, related to or associated with S. *pneumoniae*. In so far another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal, which comprises inoculating the individual with the hyperimmune serum reactive antigens and fragments thereof of the invention, or a fragment or variant thereof, adequate to produce antibodies to protect said individual from infection, particularly *streptococcal* infection and most particularly *S. pneumoniae* infections.

Yet another aspect of the invention relates to a method of inducing an immunological response in an individual which comprises, through gene therapy or otherwise, delivering a nucleic acid functionally encoding hyperimmune serum reactive antigens and fragments thereof, or a fragment or a variant thereof, for expressing the hyperimmune serum reactive antigens and fragments thereof, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibodies or a cell mediated T cell response, either cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise.

A further aspect of the invention relates to an immunological composition which, when introduced into a host capable of having induced within it an immunological response, induces an immunological response in such host, wherein the composition comprises recombinant DNA which codes for and expresses an antigen of the hyperimmune serum reactive antigens and fragments thereof of the present invention. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

The hyperimmune serum reactive antigens and fragments thereof of the invention or a fragment thereof may be fused with a co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. This fused recombinant protein preferably further comprises an antigenic co-protein, such as Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilise the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Also, provided by this invention are methods using the described nucleic acid molecule or particular

- 38 -

fragments thereof in such genetic immunization experiments in animal models of infection with *S. pneumoniae*. Such fragments will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. This approach can allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of *S. pneumoniae* infection in mammals, particularly humans.

The hyperimmune serum reactive antigens and fragments thereof may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue and mucosal tissues caused e.g. by viral infection (esp. respiratory, such as the flu) mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The present invention also includes a vaccine formulation, which comprises the immunogenic recombinant protein together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, intradermal intranasal or tramsdermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials, and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in-water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

According to another aspect, the present invention relates to a pharmaceutical composition comprising such a hyperimmune serum-reactive antigen or a fragment thereof as provided in the present invention for *S. pneumoniae.*. Such a pharmaceutical composition may comprise one, preferably at least two or more hyperimmune serum reactive antigens or fragments thereof against *S. pneumoniae*. Optionally, such *S. pneumoniae* hyperimmune serum reactive antigens or fragments thereof may also be combined with antigens against other pathogens in a combination pharmaceutical composition. Preferably, said pharmaceutical composition is a vaccine for preventing or treating an infection caused by *S. pneumoniae* and/or other pathogens against which the antigens have been included in the vaccine.

According to a further aspect, the present invention relates to a pharmaceutical composition comprising a nucleic acid molecule encoding a hyperimmune serum-reactive antigen or a fragment thereof as identified above for *S. pneumoniae*. Such a pharmaceutical composition may comprise one or more nucleic acid molecules encoding hyperimmune serum reactive antigens or fragments thereof against *S. pneumoniae*. Optionally, such *S. pneumoniae* nucleic acid molecules encoding hyperimmune serum reactive antigens or fragments thereof may also be combined with nucleic acid molecules encoding antigens against other pathogens in a combination pharmaceutical composition. Preferably, said pharmaceutical composition is a vaccine for preventing or treating an infection caused by *S. pneumoniae* and/or other pathogens against which the antigens have been included in the vaccine.

The pharmaceutical composition may contain any suitable auxiliary substances, such as buffer substances, stabilisers or further active ingredients, especially ingredients known in connection of pharmaceutical composition and/or vaccine production.

A preferable carrier/or excipient for the hyperimmune serum-reactive antigens, fragments thereof or a

coding nucleic acid molecule thereof according to the present invention is an immunostimulatory compound for further stimulating the immune response to the given hyperimmune serum-reactive antigen, fragment thereof or a coding nucleic acid molecule thereof. Preferably the immunostimulatory compound in the pharmaceutical preparation according to the present invention is selected from the group of polycationic substances, especially polycationic peptides, immunostimulatory nucleic acids molecules, preferably immunostimulatory deoxynucleotides, alum, Freund's complete adjuvants, Freund's incomplete adjuvants, neuroactive compounds, especially human growth hormone, or combinations thereof.

It is also within the scope of the present invention that the pharmaceutical composition, especially vaccine, comprises apart from the hyperimmune serum reactive antigens, fragments thereof and/or coding nucleic acid molecules thereof according to the present invention other compounds which are biologically or pharmaceutically active. Preferably, the vaccine composition comprises at least one polycationic peptide. The polycationic compound(s) to be used according to the present invention may be any polycationic compound, which shows the characteristic effects according to the WO 97/30721. Preferred polycationic compounds are selected from basic polyppetides, organic polycations, basic polyamino acids or mixtures thereof. These polyamino acids should have a chain length of at least 4 amino acid residues (WO 97/30721). Especially preferred are substances like polylysine, polyarginine and polypeptides containing more than 20 %, especially more than 50 % of basic amino acids in a range of more than 8, especially more than 20, amino acid residues or mixtures thereof. Other preferred polycations and their pharmaceutical compositions are described in WO 97/30721 (e.g. polyethyleneimine) and WO 99/38528. Preferably these polypeptides contain between 20 and 500 amino acid residues, especially between 30 and 200 residues.

These polycationic compounds may be produced chemically or recombinantly or may be derived from natural sources.

Cationic (poly)peptides may also be anti-microbial with properties as reviewed in {Ganz, T., 1999}. These (poly)peptides may be of prokaryotic or animal or plant origin or may be produced chemically or recombinantly (WO 02/13857). Peptides may also belong to the class of defensins (WO 02/13857). Sequences of such peptides can be, for example, found in the Antimicrobial Sequences Database under the following internet address:

http://www.bbcm.univ.trieste.it/~tossi/pag2.html

Such host defence peptides or defensives are also a preferred form of the polycationic polymer according to the present invention. Generally, a compound allowing as an end product activation (or down-regulation) of the adaptive immune system, preferably mediated by APCs (including dendritic cells) is used as polycationic polymer.

Especially preferred for use as polycationic substances in the present invention are cathelicidin derived antimicrobial peptides or derivatives thereof (International patent application WO 02/13857, incorporated herein by reference), especially antimicrobial peptides derived from mammalian cathelicidin, preferably from human, bovine or mouse.

Polycationic compounds derived from natural sources include HIV-REV or HIV-TAT (derived cationic peptides, antennapedia peptides, chitosan or other derivatives of chitin) or other peptides derived from these peptides or proteins by biochemical or recombinant production. Other preferred polycationic compounds are cathelin or related or derived substances from cathelin. For example, mouse cathelin is a peptide which has the amino acid sequence NH2-RLAGLLRKGGEKIGEKLKKIGOKIKNFFQKLVPQPE-COOH. Related or derived cathelin substances contain the whole or parts of the cathelin sequence with at least 15-20 amino acid residues. Derivations may include the substitution or modification of the natural

amino acids by amino acids which are not among the 20 standard amino acids. Moreover, further cationic residues may be introduced into such cathelin molecules. These cathelin molecules are preferred to be combined with the antigen. These cathelin molecules surprisingly have turned out to be also effective as an adjuvant for an antigen without the addition of further adjuvants. It is therefore possible to use such cathelin molecules as efficient adjuvants in vaccine formulations with or without further immunactivating substances.

Another preferred polycationic substance to be used according to the present invention is a synthetic peptide containing at least 2 KLK-motifs separated by a linker of 3 to 7 hydrophobic amino acids (International patent application WO 02/32451, incorporated herein by reference).

The pharmaceutical composition of the present invention may further comprise immunostimulatory nucleic acid(s). Immunostimulatory nucleic acids are e. g. neutral or artificial CpG containing nucleic acids, short stretches of nucleic acids derived from non-vertebrates or in form of short oligonucleotides (ODNs) containing non-methylated cytosine-guanine di-nucleotides (CpG) in a certain base context (e.g. described in WO 96/02555). Alternatively, also nucleic acids based on inosine and cytidine as e.g. described in the WO 01/93903, or deoxynucleic acids containing deoxy-inosine and/or deoxyuridine residues (described in WO 01/93905 and PCT/EP 02/05448, incorporated herein by reference) may preferably be used as immunostimulatory nucleic acids for the present invention. Preferablly, the mixtures of different immunostimulatory nucleic acids may be used according to the present invention.

It is also within the present invention that any of the aforementioned polycationic compounds is combined with any of the immunostimulatory nucleic acids as aforementioned. Preferably, such combinations are according to the ones as described in WO 01/93905, WO 02/32451, WO 01/54720, WO 01/93903, WO 02/13857 and PCT/EP 02/05448 and the Austrian patent application A 1924/2001, incorporated.herein by reference.

In addition or alternatively such vaccine composition may comprise apart from the hyperimmune serum reactive antigens and fragments thereof, and the coding nucleic acid molecules thereof according to the present invention a neuroactive compound. Preferably, the neuroactive compound is human growth factor as, e.g. described in WO 01/24822. Also preferably, the neuroactive compound is combined with any of the polycationic compounds and/or immunostimulatory nucleic acids as afore-mentioned.

In a further aspect the present invention is related to a pharmaceutical composition. Such pharmaceutical composition is, for example, the vaccine described herein. Also a pharmaceutical composition is a pharmaceutical composition which comprises any of the following compounds or combinations thereof: the nucleic acid molecules according to the present invention, the hyperimmune serum reactive antigens and fragments thereof according to the present invention, the vector according to the present invention, the cells according to the present invention, the antibody according to the present invention, the functional nucleic acids according to the present invention and the binding peptides such as the anticalines according to the present invention, any agonists and antagonists screened as described herein. In connection therewith any of these compounds may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a hyperimmune serum reactive antigen and fragments thereof of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal, intratracheal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application, for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1 % to about 98 % by weight of the formulation; more usually they will constitute up to about 80 % by weight of the formulation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.05-5 µg antigen / per kg of body weight, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks.

With the indicated dose range, no adverse toxicological effects should be observed with the compounds of the invention, which would preclude their administration to suitable individuals.

In a further embodiment the present invention relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. The ingredient(s) can be present in a useful amount, dosage, formulation or combination. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

In connection with the present invention any disease related use as disclosed herein such as, e. g. use of the pharmaceutical composition or vaccine, is particularly a disease or diseased condition which is caused by, linked or associated with Streptococci, more preferably, S. pneumoniae. In connection therewith it is to be noted that S. pneumoniae comprises several strains including those disclosed herein. A disease related, caused or associated with the bacterial infection to be prevented and/or treated according to the present invention includes besides others bacterial pharyngitis, otitis media, pneumonia, bacteremia, meningitis, peritonitis and sepsis in humans.

In a still further embodiment the present invention is related to a screening method using any of the hyperimmune serum reactive antigens or nucleic acids according to the present invention. Screening methods as such are known to the one skilled in the art and can be designed such that an agonist or an antagonist is screened. Preferably an antagonist is screened which in the present case inhibits or prevents the binding of any hyperimmune serum reactive antigen and fragment thereof according to the present invention to an interaction partner. Such interaction partner can be a naturally occurring interaction partner or a non-naturally occurring interaction partner.

The invention also provides a method of screening compounds to identify those, which enhance (agonist) or block (antagonist) the function of hyperimmune serum reactive antigens and fragments thereof or nucleic acid molecules of the present invention, such as its interaction with a binding molecule. The

method of screening may involve high-throughput.

For example, to screen for agonists or antagonists, the interaction partner of the nucleic acid molecule and nucleic acid, respectively, according to the present invention, maybe a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, may be prepared from a cell that expresses a molecule that binds to the hyperimmune serum reactive antigens and fragments thereof of the present invention. The preparation is incubated with labelled hyperimmune serum reactive antigens and fragments thereof in the absence or the presence of a candidate molecule, which may be an agonist or antagonist. The ability of the candidate molecule to bind the binding molecule is reflected in decreased binding of the labelled ligand. Molecules which bind gratuitously, i. e., without inducing the functional effects of the hyperimmune serum reactive antigens and fragments thereof, are most likely to be good antagonists. Molecules that bind well and elicit functional effects that are the same as or closely related to the hyperimmune serum reactive antigens and fragments thereof are good agonists.

The functional effects of potential agonists and antagonists may be measured, for instance, by determining the activity of a reporter system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect with that of the hyperimmune serum reactive antigens and fragments thereof of the present invention or molecules that elicit the same effects as the hyperimmune serum reactive antigens and fragments thereof. Reporter systems that may be useful in this regard include but are not limited to colorimetric labelled substrate converted into product, a reporter gene that is responsive to changes in the functional activity of the hyperimmune serum reactive antigens and fragments thereof, and binding assays known in the art.

Another example of an assay for antagonists is a competitive assay that combines the hyperimmune serum reactive antigens and fragments thereof of the present invention and a potential antagonist with membrane-bound binding molecules, recombinant binding molecules, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. The hyperimmune serum reactive antigens and fragments thereof can be labelled such as by radioactivity or a colorimetric compound, such that the molecule number of hyperimmune serum reactive antigens and fragments thereof bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a hyperimmune serum reactive antigen and fragments thereof of the invention and thereby inhibit or extinguish its acitivity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds to the same sites on a binding molecule without inducing functional activity of the hyperimmune serum reactive antigens and fragments thereof of the invention.

Potential antagonists include a small molecule, which binds to and occupies the binding site of the hyperimmune serum reactive antigens and fragments thereof thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

Other potential antagonists include antisense molecules (see {Okano, H. et al., 1991}; OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION; CRC Press, Boca Ration, FL (1988), for a description of these molecules).

Preferred potential antagonists include derivatives of the hyperimmune serum reactive antigens and fragments thereof of the invention.

As used herein the activity of a hyperimmune serum reactive antigen and fragment thereof according to the present invention is its capability to bind to any of its interaction partner or the extent of such capability to bind to its or any interaction partner.

In a particular aspect, the invention provides the use of the hyperimmune serum reactive antigens and fragments thereof, nucleic acid molecules or inhibitors of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: i) in the prevention of adhesion of S. pneumoniae to mammalian extracellular matrix proteins at mucosal surfaces and on in-dwelling devices or to extracellular matrix proteins in wounds; ii) to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins which mediate tissue damage or invasion iii) or lead to evasion of immune defense; iv) to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques, e.g. through inhibiting nutrient acquisition {Brown, J. et al., 2001}.

Each of the DNA coding sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein upon expression can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The antagonists and agonists may be employed, for instance, to inhibit diseases arising from infection with Streptococcus, especially S. pneumoniae, such as sepsis.

In a still further aspect the present invention is related to an affinity device such affinity device comprises as least a support material and any of the hyperimmune serum reactive antigens and fragments thereof according to the present invention, which is attached to the support material. Because of the specificity of the hyperimmune serum reactive antigens and fragments thereof according to the present invention for their target cells or target molecules or their interaction partners, the hyperimmune serum reactive antigens and fragments thereof allow a selective removal of their interaction partner(s) from any kind of sample applied to the support material provided that the conditions for binding are met. The sample may be a biological or medical sample, including but not limited to, fermentation broth, cell debris, cell preparation, tissue preparation, organ preparation, blood, urine, lymph liquid, liquor and the like.

The hyperimmune serum reactive antigens and fragments thereof may be attached to the matrix in a covalent or non-covalent manner. Suitable support material is known to the one skilled in the art and can be selected from the group comprising cellulose, silicon, glass, aluminium, paramagnetic beads, starch and dextrane.

The present invention is further illustrated by the following figures, examples and the sequence listing, from which further features, embodiments and advantages may be taken. It is to be understood that the present examples are given by way of illustration only and not by way of limitation of the disclosure.

In connection with the present invention

Figure 1 shows the characterization of S. pneumoniae specific human sera.

Figure 2 shows the characterization of the small fragment genomic library, LSPn-70, from Streptococcus pneumoniae serotype 4.

Figure 3 shows the selection of bacterial cells by MACS using biotinylated human IgGs.

- 44 -

Figure 4 shows an example for the gene distribution study with the identified antigens.

Figure 5 shows examples of changes in epitope-specific antibody levels in the different age groups and during the course of pneumococcal disease.

Figure 6 shows examples for cell surface staining with epitope-specific antisera by flow cytometry.

Figure 7 shows the determination of bactericidal activity of antibodies induced by selected epitopes in an *in vitro assay*.

Figure 8 shows the protective effect of active immunization with selected *S. pneumoniae* antigens in a murine lethality sepsis model.

Figure 9 shows the protective effect of passive immunization with sera generated with selected *S. pneumoniae* antigens in a murine lethality sepsis model.

Figure 10 shows the identification of the protective domain within the SP2216 antigen.

Figure 11 shows that antibodies induced by protective antigens are cross-reactive with the different *S. pneumoniae* serotypes.

Figure 12 shows the alignment of amino acid sequences of natural SP2216 variants

Figure 13 shows the alignment of amino acid sequences of natural SP1732 variants

Figure 14 shows the alignment of amino acid sequences of natural SP2190 variants

Table 1 shows the summary of all screens performed with genomic *S. pneumoniae* libraries and human serum.

Table 2 shows the summary of epitope serology analysis with human sera.

Table 3 shows the summary of the gene distribution analysis for the identified antigens in 50 *S. pneumoniae* strains.

Table 4 shows the summary of the surface staining and bactericidal activity measurements.

The figures to which it might be referred to in the specification are described in the following in more details.

Figure 1 shows the characterization of human sera for anti-*S. pneumoniae* antibodies ass measured by immune assays. Total anti-*S. pneumoniae* IgG and IgA antibody levels were measured by standard ELISA using total bacterial lysates or culture supernant fractions prepared from *S. pneumoniae* serotype 4 capsule negative mutant strain as coating antigens. 97 serum samples from convalescing patients with inasive diseases or 50 sera from healthy adults without nasopharyngeal carriage of *S. pneumoniae* were analysed at three different serum dilutions. Results of representative experiments are shown with (A) patients' sera with bacterial lysate and (B) healthy adult sera with culture supermtant proteins. Data are expressed as ELISA units calculated from absorbance at 405nm at a serum dilution in the linear range of detection (10.000X for IgA, 50,000 for IgG). 2x5 sera from both donor groups were selected and pooled for antigen identification by bacterial surface display. Selected sera included in the two patient (PSPn3-IgG,-IgA and PSPn7-IgG) and two healthy pools (NSPn4-IgG,-IgA and NSPn5-IgG) are indicated by circules. (C) Immunoblot analysis was performed on sera pre-selected by ELISA in order to ensure multiple immune

- 45 -

reactivity with protein antigens. Results of a representative experiment using total bacterial lysate prepared from *S. pneumoniae* serotype 4 capsule negative mutant strain and selected patients' sera at 5.000X dilution are shown. Not selected, low titer sera were included as negative controls. Mw: molecular weigth markers. **(D)** Surface staining of *S. pneumoniae* serotype 4 capsule negative mutant strain was performed by FACS to compare antibody binding to surface located antigens. Human sera were used at different concentrations (0.5-5%). Representative data are shown with patients' sera used at 0.5% final concentration. Signal was detected with FITC-labeled anti-human IgGFab and analysed with the computer program CELLQuest. **(E)** Correlation between IgG titers measured by ELISA using total bacterial lysates and surface staining of whole living *S. pneumoniae* with serum IgGs is shown. IgG titer is expressed as ELISA units, while surface staining is expressed as mean fluorescence of stained bacteria calculated by the computer program CELLQuest .

Figure 2 (A) shows the fragment size distribution of the *Streptococcus pneumoniae* type 4 small fragment genomic library, LSPn-70. After sequencing 609 randomly selected clones sequences were trimmed to eliminate vector residues and the number of clones with various genomic fragment sizes were plotted. **(B)** shows graphic illustration of the distribution of the same set of randomly sequenced clones of LSPn-70 over the *S. pneumoniae* chromosome (according to the TIGR4 genome data). Circles indicate matching sequences to annotated ORFs and rectangles represent fully matched clones to non-coding chromosomal sequences in +/+ or +/- orientation. Diamonds position all clones with chimeric sequences. Numeric distances in base pairs are indicated over the circular genome for orientation. Partitioning of various clone sets within the library is given in numbers and percentage at the bottom of the figure.

Figure 3 (A) shows the MACS selection with biotinylated human IgGs. The LSPn-70 library in pMAL9.1 was screened with 10 μg biotinylated IgG (PSPn3-IgG, purified from human serum). As negative control, no serum was added to the library cells for screening. Number of cells selected after the 1st and 2nd elution are shown for each selection round (upper and lower panel, respectively). **(B)** shows the reactivity of specific clones (1-26) selected by bacterial surface display as analysed by immunoblot analysis with the human serum IgG pool (PSPn7-IgG, 4μg/μl) used for selection by MACS at a dilution of 1:3,000. As a loading control the same blot was also analysed with antibodies directed against the platform protein LamB at a dilution of 1:5,000 of hyperimmune rabbit serum. LB, Extract from a clone expressing LamB without foreign peptide insert.

Figure 4 (A) shows the representation of different serotypes of *S. pneumoniae* clinical isolates analysed for the gene distribution study. **(B)** shows the PCR analysis for the gene distribution of SP1604 with the respective oligonucleotides. The predicted size of the PCR fragments is 470 bp. 1-50, *S. pneumoniae* strains, clinical isolates as listed under A; -, no genomic DNA added; +, genomic DNA from *S. pneumoniae* serotype 4, which served as template for library construction.

Figure 5 shows the ELISA measurement of epitope-specific human serum IgG antibody levels during pneumococcal disease. Three serum samples were collected longitudinally from patients with invasive pneumococcal disease, before disease occurred (pre), in the acute and convalescent phases. Representative experiments are shown with two sets of sera from two different patients, (A) P1147 and (B) P1150 reacted with peptides representing the identified antigens SP0069, SP0082, SP0117, SP1175, SP1937, SP2190 and SP2216, as indicated. Biotin-labeled peptides were reacted with human serum samples at 200X and 1.000X dilutions and data are expressed as ELISA units.

Figure 6 shows the detection of specific antibody binding on the cell surface of *Streptococcus pneumoniae* by flow cytometry. In Figure 5A preimmune mouse sera and polyclonal sera raised against *S. pneumoniae* serotype 4 lysate were incubated with *S. pneumoniae* strain serotype 4 and analysed by flow cytometry. Control shows the level of non-specific binding of the secondary antibody to the surface of *S. pneumoniae* cells. The histograms in figure 5B indicates the increased fluorescence due to specific binding of anti-SP2216, anti-SP0117, anti-SP0454 and anti-CRF1992 antibodies in comparison to the control sera against

the platform protein LamB.

Figure 7 shows the bactericidal activity of epitope specific antibodies as determined in *in vitro* killing assay. The killing activity of immune sera is measured parallel with and calculated relative to the appropriate control sera. Data are expressed as percentage of killing, that is the reduction on bacterial cfu numbers as a consequence of the presence of antibodies in hyperimmune (HI) polyclonal mouse sera generated with *S. pneumoniae* lysate (A), in immune sera generated with SP0117 epitopes expressed in the LamB platform protein (B), and in mouse immune sera generated with SP1287 epitopes expressed in the FhuA platform protein (C). The control sera represent preimmune sera (PI), sera induced with Lamb or FhuA expressing *E. coli* clones without *S. pneumonia*-derived epitopes. *S. pneumonaie* serotype 4 cells were incubated with mouse phagocytic cells for 60 min, and surviving bacteria were quantified by counting cfus after plating on blood agar.

Figure 8 shows the protection achieved by active immunization with selected *S. pneumoniae* antigens in a mouse lethality model. C3H mice (10 in each test groups) were immunized with recombinant antigens cloned from a serotype 4 *S. pneumoniae* strain and challenged with serotype 6B strain. Survival was monitored for 2 to 3 weeks post-challenge. A: Mice were immunized with SP0368, SP0667, SP2190 and SP2216 antigens and challenged intravenously with 10^4 cfu bacteria. Nonimmunized mice were used as negative controls, while PspA (SP0117) served as positive control. B: Mice were immunized with SP2190 and SP2216 antigens and challenged intraperitonealy with 10^5 cfu bacteria. Mice injected with PBS or mock immunized with the adjuvants only (CFA/IFA) were used as negative controls, while PspA (SP0117) served as positive control. C: Mice were immunized with SP0498 and SP1732 antigens and challenged intraperitonealy with 10^5 cfu bacteria. Mice injected with PBS were used as negative controls, while PspA (SP0117) served as positive control.

Figure 9 shows the protection achieved by passive immunization with hyperimmune mouse sera generated with selected *S. pneumoniae* antigens in a mouse lethality model. C3H mice (10 in each test groups) were given mouse sera intraperitoneally 2 hrs before intraperitoneal challenge with 10^5 cfu *S. pneumoniae* serotype 6B bacteria. Survival was monitored for 3 weeks post-challenge. 150µl immune sera generated with SP2190 or SP2216 were given and supplemented with 150µl serum from naive mice, except for mice receiving 100µl each of anti-SP2190, anti-SP2216 immune sera and 100 µl serum from naive mice. Negative controls were treated either with 300µl sera from PBS injected, noninjected (naive) or nonimmune CFA/IFA injected mice.

Figure 10 shows the identification of the protective domain within the SP2216 antigen. A: Schematic representation of the SP2216 antigen indicating the two subdomains predicted by *in silico* (structural prediction) analysis and the localization of epitopes identified by bacterial surface display (grey bars and arrows). B: C3H mice (10 in each test groups) were immunized with recombinant SP2216 antigens: full-length, N-terminal or C-terminal domains and challenged with *S. pneumoniae* serotype 6B strain given intraperitoneally 10^5 cfus. Survival was monitored for 2 to 3 weeks post-challenge. Nonimmunized (CFA/IFA adjuvant injected) mice were used as negative controls, while PspA (SP0117) served as positive control.

Figure 11 shows the cross-reactivity of antibodies by analysing different *S. pneumoniae* serotypes. Immunoblot analysis was performed with bacterial lysates prepared from 60 clinical isolates of *S. pneumoniae* representing 48 different serotypes and using sera generated with SP1732, SP2190 and SP2216 recombinant antigens cloned from a serotype 4 strain in order to test the cross-reactivity of antibodies. Results with seven different serotypes (Lanes 1-7) are shown as representative data taken from the complete analysis. Mw: molecular weight markers.

Figure 12 shows the amino acid exchanges detected in natural SP2216 variants expressed in different clinical isolates of *S. pneumoniae*. The SP2216 gene from 47 different clinical isolates representing 47 different *S. pneumoniae* serotypes were analysed by DNA sequencing. The translated amino acid

- 47 -

sequences are shown for those serotypes where amino acid exchanges were detected relative to the published TIGR4 genome sequences.

Figure 13 shows the amino acid exchanges detected in natural SP1732 variants expressed in different clinical isolates of S. pneumoniae. The SP1732 gene from six different clinical isolates representing six major S. pneumoniae serotypes (4, 6B, 9V, 18C, 19F, 23F) were analysed by DNA sequencing. The translated amino acid sequences are shown for those serotypes where amino acid exchanges were detected relative to the published TIGR4 (serotype 4) genome sequences.

Figure 14 shows the amino acid exchanges detected in natural SP2190 variants expressed in different clinical isolates of S. pneumoniae. The SP2190 gene from seven different clinical isolates representing seven major S. pneumoniae serotypes (4, 6B, 9V, 14, 18C, 19F, 23F) were analysed by DNA sequencing. The translated amino acid sequences are shown for those serotypes where amino acid exchanges were detected relative to the published TIGR4 (serotype 4) genome sequences. Due to missing sequence information from the middle part of some of the genes, the N-terminal and C-terminal aa alignments are shown separately. A: N-terminal amino acid sequences; B: C-terminal amino acid sequences.

Table 1: Immunogenic proteins identified by bacterial surface display.
A, 300bp library in fhuA with NSPn4-IgA (362), B, 300bp library in fhuA with NSPn4-IgG (832), C, 300bp library in fhuA with NSPn5-IgG (872), D, 300bp library in fhuA with PSPn3-IgA (361), E, 300bp library in fhuA with PSPn3-IgG (575), F, 300bp library in fhuA with PSPn7-IgG (795), G, 70bp library in lamB with NSPn4-IgA (1043), H, 70bp library in lamB with NSPn4-IgG (929), I, 70bp library in lamB with NSPn5-IgG (527), K, 70bp library in lamB with PSPn3-IgA (1121), L, 70bp library in lamB with PSPn3-IgG (1242), M, 70bp library in lamB with PSPn7-IgG (514); *, prediction of antigenic sequences longer than 5 amino acids was performed with the program ANTIGENIC {Kolaskar, A. et al., 1990}.

Table 2: Epitope serology with human sera.
Immune reactivity of individual synthetic peptides representing selected epitopes with individual human sera is shown. Extent of reactivity is pattern/grey coded; white: - (<50U), light grey: + (50-119U), dark gery: ++ (120-199U), black: +++ (200-500U) and vertically crossed: ++++ (> 500U). ELISA units (U) are calculated from OD_{405nm} readings and the serum dilution after correction for background. S stands for score, calculated as the sum of all reactivities (addition of the number of all +); P1 to P13 sera are measured to be high titer and are from patients with invasive penumococcal diseases and N1 to N10 sera are from healthy adults with high anti-S. pneumoniae titers. S stands for score. Which is the sum of immune reactivities: - =0; + =1; ++ =2; +++ =3 and ++++ =4. Location of synthetic peptides within the antigenic ORFs according to the genome annotation of TIGR4 strain are given in columns from and to indicating the first and last amino acid residues, respectively. Peptide names: SP0117.1-7 present in annotated ORFSP0117; ARF0408.1, potential novel ORF in alternative reading-frame of SP0408; CRF0129.1, potential novel ORF on complement of SP0129.

Table 3: Gene distribution in S. pneumoniae strains.
Fifty S. pneumoniae strains as shown in Figure 4A were tested by PCR with oligonucleotides specific for the genes encoding relevant antigens. The PCR fragment of one selected PCR fragment was sequenced in order to confirm the amplification of the correct DNA fragment. *, number of amino acid substitutions in a serotype 14 strain as compared to S. pneumoniae TIGR4 (serotype 4). #, alternative strain used for sequencing, because gene was not present in the serotype 14 strain.

Table 4: Surface location of antigenic epitopes and the functionality of the epitope-specific antibodies.
45 S. pneumoniae antigens were tested for surface localization in the way described and presented in Figure 6 by using mouse sera generated by immunization with E. coli clones harboring plasmids encoding the platform proteins LamB or FhuA fused to a S. pneumoniae peptide. Data are summarized in the column labeled FACS. The very same immune reagents were used in an in vitro killing assay, as

shown in Figure 7 for the examples, and presented for all antigens tested positive by FACS in column **PK** (phagocytic killing). -: negative result, +: not consistently positive in all assays performed, ++ and +++ are consistently positive relative to control reagents.

EXAMPLES

Example 1: Characterization and selection of human sera based anti-*S. pneumoniae* antibodies, preparation of antibody screening reagents

Experimental procedures

Enzyme linked immune assay (ELISA).
ELISA plates (Maxisorb, Millipore) were coated with 5-10 µg/ml total protein diluted in coating buffer (0.1M sodium carbonate pH 9.2). Three dilutions of sera (2,000X, 10,000X, 50,000X) were made in PBS-BSA. Highly specific Horse Radish Peroxidase (HRP)-conjugated anti-human IgG or anti-human IgA secondary antibodies (Southern Biotech) were used according to the manufacturers' recommendations (dilution: 1,000x). Antigen-antibody complexes were quantified by measuring the conversion of the substrate (ABTS) to colored product based on OD_{405nm} readings by automatic ELIAS reader (TECAN SUNRISE).

Preparation of bacterial antigen extracts
Total bacterial lysate: Bacteria were grown overnight in THB (Todd-Hewitt Broth) and lysed by repeated freeze-thaw cycles: incubation on dry ice/ethanol-mixture until frozen (1 min), then thawed at 37°C (5 min): repeated 3 times. This was followed by sonication and collection of supernatant by centrifugation (3,500 rpm, 15 min, 4°C).
Culture supernatant: After removal of bacteria by centrifugation, the supernatant of overnight grown bacterial cultures was precipitated with ice-cold ethanol by mixing 1 part supernatant with 3 parts abs. ethanol and incubated overnight at -20°C. Precipitates were collected by centrifugation (2,600 g, for 15 min). Dry pellets were dissolved either in PBS for ELISA, or in urea and SDS-sample buffer for SDS-PAGE and immunoblotting. The protein concentration of samples was determined by Bradford assay.

Immunoblotting
Total bacterial lysate and culture supernatant samples were prepared from *in vitro* grown *S. pneumoniae* serotype 4 uncapsulated mutant strain. 10 to 25µg total protein/lane was separated by SDS-PAGE using the BioRad Mini-Protean 3 Cell electrophoresis system and proteins transferred to nitrocellulose membrane (ECL, Amersham Pharmacia). After overnight blocking in 5% milk, human sera were added at 2,000x dilution, and HRPO labeled anti-human IgG was used for detection.

Surface staining of bacteria
Flow cytometric analysis was carried out as follows. *S. pneumoniae* serotype 4 uncapsulated mutant strain was grown in Todd-Hewitt broth overnight until early stationary phase. Cells were collected and washed twice in Hanks Balanced Salt Solution (HBSS) and the cell density was adjusted to approximately 1×10^6 CFU in 100µl HBSS with 0.5% BSA based on OD600 nm readings. After incubation with human sera at 0,5 and 2% final concentration for 60 min at 4°C, unbound antibodies were washed away by centrifugation in excess HBSS, 0.5% BSA. For detection fluorescein (FITC) labeled secondary goat anti-human IgG (F(ab')2 fragment specific) was incubated with the cells at 4°C for 30 min. After washing the cells, cells were fixed with 2% paraformaldehyde. Surface staining antibodies were detected using a Becton Dickinson FACScan flow cytometer and data further analyzed with the computer program CELLQuest.

Purification of antibodies for genomic screening. Five sera from both the patient and the healthy group were selected based on the overall anti-streptococcal titers for a serum pool used in the screening procedure.

- 49 -

Antibodies against *E. coli* proteins were removed by incubating the heat-inactivated sera with whole cell *E. coli* cells (DH5alpha, transformed with pHIE11, grown under the same condition as used for bacterial surface display). Highly enriched preparations of IgGs from the pooled, depleted sera were generated by protein G affinity chromatography, according to the manufacturer's instructions (UltraLink Immobilized Protein G, Pierce). IgA antibodies were purified also by affinity chromatography using biotin-labeled anti-human IgA (Southern Biotech) immobilized on Streptavidin-agarose (GIBCO BRL). The efficiency of depletion and purification was checked by SDS-PAGE, Western blotting, ELISA and protein concentration measurements.

Results

The antibodies produced against *S. pneumoniae* by the human immune system and present in human sera are indicative of the *in vivo* expression of the antigenic proteins and their immunogenicity. These molecules are essential for the identification of individual antigens in the approach as described in the present invention, which is based on the interaction of the specific anti-streptococcal antibodies and the corresponding *S. pneumoniae* peptides or proteins. To gain access to relevant antibody repertoires, human sera were collected from

       I. convalescent patients with invasive *S. pneumoniae* infections, such as pneumonia, bacteraemia and meningitis. (*S. pneumoniae* was shown to be the causative agent by medical microbiological tests),

       II healthy adults without carriage at the time of sampling. *S. pneumoniae* colonization and infections are common, and antibodies are present as a consequence of natural immunization from previous encounters.

97 serum samples from patien and 50 sera from healthy adults were characterized for anti-*S. pneumoniae* antibodies by a series of immune assays. Primary characterization was done by ELISA using two different antigen preparations, such as total bacterial extract and culture supernatant proteins prepared from *S. pneumoniae* serotype 4 uncapsulated mutant strain. It is an important aspect that we analysed uncapsulated strain, since we avoided the reactivities coming from serotype specific abundant anti-capsular polysaccharide antibodies.

Recently it was reported that not only IgG, but also IgA serum antibodies can be recognized by the FcRIII receptors of PMNs and promote opsonization {Phillips-Quagliata, J. et al., 2000}; {Shibuya, A. et al., 2000}. The primary role of IgA antibodies is neutralization, mainly at the mucosal surface. The level of serum IgA reflects the quality, quantity and specificity of the dimeric secretory IgA. For that reason the serum collection was not only analyzed for anti-streptococcal IgG, but also for IgA levels. In the ELISA assays highly specific secondary reagents were used to detect antibodies from the high affinity types, such as IgG and IgA, but avoided IgM. Production of IgM antibodies occurs during the primary adaptive humoral response, and results in low affinity antibodies, while IgG and IgA antibodies had already undergone affinity maturation, and are more valuable in fighting or preventing disease. Antibody titers were compared at given dilutions where the response was linear (Fig. 1A and 1B.). Sera were ranked based on the IgG and IgA reactivity against the two complex antigenic mixtures, and the highest ones were selected for further testing by immunoblotting. This analysis confirmed a high antibody reactivity of the pre-selected sera against multiple pneumococcal proteins, especially when compared to not selected, low-titer sera (Fig 1C). ELISA ranking of sera also correlated very well with surface staining of the same *S. pneumoniae* strain (Fig. 1D and 1E) suggesting that the majority of the antibodies detected by ELISA corresponded to surface antigens. This extensive antibody characterization approach has led to the unambiguous identification of anti-pneumococcal hyperimmune sera.

Selected sera, 2x5 from both the patient and healthy donor groups were pooled to further enrich for abundant antibodies, but still having a representation of antibody repertoires of different individuals. IgG and IgA antibodies were purified from pooled sera by affinity chromatography and depleted of E. coli -reactive antibodies to avoid background in the bacterial surface display screen.

- 50 -

Example 2: Generation of highly random, frame-selected, small-fragment, genomic DNA libraries of *Streptococcus pneumoniae*

Experimental procedures

Preparation of streptococcal genomic DNA. 50 ml Todd-Hewitt Broth medium was inoculated with *S. pneumoniae* serotype 4 (clinical isolate, typed with conventional serotyping) bacteria from a frozen stab and grown with aeration and shaking for 18 h at 37°C. The culture was then harvested, centrifuged with 1,600x g for 15 min and the supernatant was removed. Bacterial pellets were washed 3 x with PBS and carefully re-suspended in 0.5 ml of Lysozyme solution (100 mg/ml). 0.1 ml of 10 mg/ml heat treated RNase A and 20 U of RNase T1 were added, mixed carefully and the solution was incubated for 1 h at 37°C. Following the addition of 0.2 ml of 20 % SDS solution and 0.1 ml of Proteinase K (10 mg/ml) the tube was incubated overnight at 55°C. 1/3 volume of saturated NaCl was then added and the solution was incubated for 20 min at 4°C. The extract was pelleted in a microfuge (13,000 rpm) and the supernatant transferred into a new tube. The solution was extracted with PhOH/CHCl₃/IAA (25:24:1) and with CHCl₃/IAA (24:1). DNA was precipitated at room temperature by adding 0.6x volume of Isopropanol, spooled from the solution with a sterile Pasteur pipette and transferred into tubes containing 80% ice-cold ethanol. DNA was recovered by centrifuging the precipitates with 10-12,000x g, then dried on air and dissolved in ddH_2O.

Preparation of small genomic DNA fragments. Genomic DNA fragments were mechanically sheared into fragments ranging in size between 150 and 300 bp using a cup-horn sonicator (Bandelin Sonoplus UV 2200 sonicator equipped with a BB5 cup horn, 10 sec. pulses at 100 % power output) or into fragments of size between 50 and 70 bp by mild DNase I treatment (Novagen). It was observed that sonication yielded a much tighter fragment size distribution when breaking the DNA into fragments of the 150-300 bp size range. However, despite extensive exposure of the DNA to ultrasonic wave-induced hydromechanical shearing force, subsequent decrease in fragment size could not be efficiently and reproducibly achieved. Therefore, fragments of 50 to 70 bp in size were obtained by mild DNase I treatment using Novagen's shotgun cleavage kit. A 1:20 dilution of DNase I provided with the kit was prepared and the digestion was performed in the presence of MnCl₂ in a 60 µl volume at 20°C for 5 min to ensure double-stranded cleavage by the enzyme. Reactions were stopped with 2 µl of 0.5 M EDTA and the fragmentation efficiency was evaluated on a 2% TAE-agarose gel. This treatment resulted in total fragmentation of genomic DNA into near 50-70 bp fragments. Fragments were then blunt-ended twice using T4 DNA Polymerase in the presence of 100 µM each of dNTPs to ensure efficient flushing of the ends. Fragments were used immediately in ligation reactions or frozen at -20°C for subsequent use.

Description of the vectors. The vector pMAL4.31 was constructed on a pASK-IBA backbone {Skerra, A., 1994} with the beta-lactamase (*bla*) gene exchanged with the Kanamycin resistance gene. In addition the *bla* gene was cloned into the multiple cloning site. The sequence encoding mature beta-lactamase is preceded by the leader peptide sequence of *ompA* to allow efficient secretion across the cytoplasmic membrane. Furthermore a sequence encoding the first 12 amino acids (spacer sequence) of mature beta-lactamase follows the *ompA* leader peptide sequence to avoid fusion of sequences immediately after the leader peptidase cleavage site, since e.g. clusters of positive charged amino acids in this region would decrease or abolish translocation across the cytoplasmic membrane {Kajava, A. et al., 2000}. A *SmaI* restriction site serves for library insertion. An upstream FseI site and a downstream NotI site, which were used for recovery of the selected fragment, flank the *SmaI* site. The three restriction sites are inserted after the sequence encoding the 12 amino acid spacer sequence in such a way that the *bla* gene is transcribed in the -1 reading frame resulting in a stop codon 15 bp after the *NotI* site. A +1 bp insertion restores the *bla* ORF so that beta-lactamase protein is produced with a consequent gain of Ampicillin resistance.

The vector pMAL9.1 was constructed by cloning the *lamB* gene into the multiple cloning site of pEH1 {Hashemzadeh-Bonehi, L. et al., 1998}. Subsequently, a sequence was inserted in *lamB* after amino acid

- 51 -

154, containing the restriction sites *FseI*, *SmaI* and *NotI*. The reading frame for this insertion was constructed in such a way that transfer of frame-selected DNA fragments excised by digestion with *FseI* and *NotI* from plasmid pMAL4.31 yields a continuous reading frame of *lamB* and the respective insert.

The vector pMAL10.1 was constructed by cloning the *btuB* gene into the multiple cloning site of pEH1. Subsequently, a sequence was inserted in *btuB* after amino acid 236, containing the restriction sites *FseI*, *XbaI* and *NotI*. The reading frame for this insertion was chosen in a way that transfer of frame-selected DNA fragments excised by digestion with *FseI* and *NotI* from plasmid pMAL4.31 yields a continuous reading frame of *btuB* and the respective insert.

The vector pHIE11 was constructed by cloning the *fhuA* gene into the multiple cloning site of pEH1. Thereafter, a sequence was inserted in *fhuA* after amino acid 405, containing the restriction site *FseI*, *XbaI* and *NotI*. The reading frame for this insertion was chosen in a way that transfer of frame-selected DNA fragments excised by digestion with *FseI* and *NotI* from plasmid pMAL4.31 yields a continuous reading frame of *fhuA* and the respective insert.

Cloning and evaluation of the library for frame selection. Genomic *S. pneumoniae* DNA fragments were ligated into the *SmaI* site of the vector pMAL4.31. Recombinant DNA was electroporated into DH10B electrocompetent *E. coli* cells (GIBCO BRL) and transformants plated on LB-agar supplemented with Kanamycin (50 µg/ml) and Ampicillin (50 µg/ml). Plates were incubated over night at 37°C and colonies collected for large scale DNA extraction. A representative plate was stored and saved for collecting colonies for colony PCR analysis and large-scale sequencing. A simple colony PCR assay was used to initially determine the rough fragment size distribution as well as insertion efficiency. From sequencing data the precise fragment size was evaluated, junction intactness at the insertion site as well as the frame selection accuracy (*3n+1 rule*).

Cloning and evaluation of the library for bacterial surface display. Genomic DNA fragments were excised from the pMAL4.31 vector, containing the *S. pneumoniae* library with the restriction enzymes *FseI* and *NotI*. The entire population of fragments was then transferred into plasmids pMAL9.1 (LamB) or pHIE11 (FhuA), which have been digested with *FseI* and *NotI*. Using these two restriction enzymes, which recognise an 8 bp GC rich sequence, the reading frame that was selected in the pMAL4.31 vector is maintained in each of the platform vectors. The plasmid library was then transformed into *E. coli* DH5alpha cells by electroporation. Cells were plated onto large LB-agar plates supplemented with 50 µg/ml Kanamycin and grown over night at 37°C at a density yielding clearly visible single colonies. Cells were then scraped off the surface of these plates, washed with fresh LB medium and stored in aliquots for library screening at -80°C.

Results

Libraries for frame selection. Two libraries (LSPn70 and LSPn300) were generated in the pMAL4.31 vector with sizes of approximately 70 and 300 bp, respectively. For each library, ligation and subsequent transformation of approximately 1 µg of pMAL4.31 plasmid DNA and 50 ng of fragmented genomic *S. pneumoniae* DNA yielded 4×10^5 to 2×10^6 clones after frame selection. To assess the randomness of the libraries, approximately 600 randomly chosen clones of LSPn70 were sequenced. The bioinformatic analysis showed that of these clones only very few were present more than once. Furthermore, it was shown that 90% of the clones fell in the size range between 25 and 100 bp with an average size of 52 bp (Figure 2). Allmost all sequences followed the 3n+1 rule, showing that all clones were properly frame selected.

Bacterial surface display libraries. The display of peptides on the surface of *E. coli* required the transfer of the inserts from the LSPn libraries from the frame selection vector pMAL4.31 to the display plasmids pMAL9.1 (LamB) or pHIE11 (FhuA). Genomic DNA fragments were excised by *FseI* and *NotI* restriction

and ligation of 5ng inserts with 0.1µg plasmid DNA and subsequent transformation into DH5alpha cells resulted in 2-5x 10^6 clones. The clones were scraped off the LB plates and frozen without further amplification.

Example 3: Identification of highly immunogenic peptide sequences from *S. pneumoniae* using bacterial surface displayed genomic libraries and human serum

Experimental procedures

MACS screening. Approximately 2.5x 10^8 cells from a given library were grown in 5 ml LB-medium supplemented with 50 µg/ml Kanamycin for 2 h at 37°C. Expression was induced by the addition of 1 mM IPTG for 30 min. Cells were washed twice with fresh LB medium and approximately 2x 10^7 cells re-suspended in 100 µl LB medium and transferred to an Eppendorf tube.

10 µg of biotinylated, human IgGs purified from serum was added to the cells and the suspension incubated overnight at 4°C with gentle shaking. 900 µl of LB medium was added, the suspension mixed and subsequently centrifuged for 10 min at 6,000 rpm at 4°C (For IgA screens, 10 µg of purified IgAs were used and these captured with biotinylated anti-human-IgG secondary antibodies). Cells were washed once with 1 ml LB and then re-suspended in 100 µl LB medium. 10 µl of MACS microbeads coupled to streptavidin (Miltenyi Biotech, Germany) were added and the incubation continued for 20 min at 4°C. Thereafter 900 µl of LB medium was added and the MACS microbead cell suspension was loaded onto the equilibrated MS column (Miltenyi Biotech, Germany) which was fixed to the magnet. (The MS columns were equilibrated by washing once with 1 ml 70% EtOH and twice with 2 ml LB medium.)

The column was then washed three times with 3 ml LB medium. After removal of the magnet, cells were eluted by washing with 2 ml LB medium. After washing the column with 3 ml LB medium, the 2 ml eluate was loaded a second time on the same column and the washing and elution process repeated. The loading, washing and elution process was performed a third time, resulting in a final eluate of 2 ml.

A second round of screening was performed as follows. The cells from the final eluate were collected by centrifugation and re-suspended in 1 ml LB medium supplemented with 50 µg/ml Kanamycin. The culture was incubated at 37°C for 90 min and then induced with 1 mM IPTG for 30 min. Cells were subsequently collected, washed once with 1 ml LB medium and suspended in 10 µl LB medium. 10 µg of human, biotinylated IgGs were added again and the suspension incubated over night at 4°C with gentle shaking. All further steps were exactly the same as in the first selection round. Cells selected after two rounds of selection were plated onto LB-agar plates supplemented with 50 µg/ml Kanamycin and grown over night at 37°C.

Evaluation of selected clones by sequencing and Western blot analysis. Selected clones were grown overnight at 37°C in 3 ml LB medium supplemented with 50 µg/ml Kanamycin to prepare plasmid DNA using standard procedures. Sequencing was performed at MWG (Germany) or in collaboration with TIGR (U.S.A.).

For Western blot analysis approximately 10 to 20 µg of total cellular protein was separated by 10% SDS-PAGE and blotted onto HybondC membrane (Amersham Pharmacia Biotech, England). The LamB or FhuA fusion proteins were detected using human serum as the primary antibody at a dilution of approximately 1:5,000 and anti-human IgG or IgA antibodies coupled to HRP at a dilution of 1:5,000 as secondary antibodies. Detection was performed using the ECL detection kit (Amersham Pharmacia Biotech, England). Alternatively, rabbit anti-FhuA or rabbit anti-LamB polyclonal immune sera were used as primary antibodies in combination with the respective secondary antibodies coupled to HRP for the detection of the fusion proteins.

Results

Screening of bacterial surface display libraries by magnetic activated cell sorting (MACS) using biotinylated Igs.
The libraries LSPn70 in pMAL9.1 and LSPn300 in pHIE11 were screened with pools of biotinylated, human IgGs and IgAs from patient sera or sera from healthy individuals (see Example 1: *Preparation of antibodies from human serum*). The selection procedure was performed as described under Experimental procedures. Figure 3A shows a representative example of a screen with the LSPn-70 library and PSPn3-IgGs. As can be seen from the colony count after the first selection cycle from MACS screening, the total number of cells recovered at the end is drastically reduced from 2x10⁷ cells to approximately 5x 10⁴ cells, whereas the selection without antibodies added showed a reduction to about 2x10³ cells (Figure 3A). After the second round, a similar number of cells was recovered with PSPn3-IgGs, while fewer than 10 cells were recovered when no IgGs from human serum were added, clearly showing that selection was dependent on *S. pneumoniae* specific antibodies. To evaluate the performance of the screen, 26 selected clones were picked randomly and subjected to immunoblot analysis with screening IgG pool (PSPn7) (Figure 3B). This analysis revealed that ~90% of the selected clones showed reactivity with antibodies present in the relevant serum whereas the control strain expressing LamB without a *S. pneumoniae* specific insert did not react with the same serum. In general, the rate of reactivity was observed to lie within the range of 35 to 90%. Colony PCR analysis showed that all selected clones contained an insert in the expected size range.

Subsequent sequencing of a larger number of randomly picked clones (600 to 1200 per screen) led to the identification of the gene and the corresponding peptide or protein sequence that was specifically recognized by the human serum antibodies used for screening. The frequency with which a specific clone is selected reflects at least in part the abundance and/or affinity of the specific antibodies in the serum used for selection and recognizing the epitope presented by this clone. In that regard it is striking that clones derived from some ORFs (e.g. SP2216, SP0117, SP0641, SP2136, SP2190, SP0107, SP0082) were picked more than 100 times, indicating their highly immunogenic property. Table 1 summarizes the data obtained for all 12 performed screens. All clones that are presented in Table 1 have been verified by immunoblot analysis using whole cellular extracts from single clones to show the indicated reactivity with the pool of human serum used in the respective screen. As can be seen from Table 1, distinct regions of the identified ORF are identified as immunogenic, since variably sized fragments of the proteins are displayed on the surface by the platform proteins.

It is further worth noticing that most of the genes identified by the bacterial surface display screen encode proteins that are either attached to the surface of *S. pneumoniae* and/or are secreted. This is in accordance with the expected role of surface attached or secreted proteins in virulence of *S. pneumoniae*.

Example 4: Assessment of the reactivity of highly immunogenic peptide sequences with individual human sera.

Experimental procedures

Peptide synthesis
Peptides were synthesized in small scale (4 mg resin; up to 288 in parallel) using standard F-moc chemistry on a Rink amide resin (PepChem, Tübingen, Germany) using a SyroII synthesizer (Multisyntech, Witten, Germany). After the sequence was assembled, peptides were elongated with Fmoc-epsilon-aminohexanoic acid (as a linker) and biotin (Sigma, St. Louis, MO; activated like a normal amino acid). Peptides were cleaved off the resin with 93%TFA, 5% triethylsilane, and 2% water for one hour. Peptides were dried under vacuum and freeze dried three times from acetonitrile/water (1:1). The presence of the correct mass was verified by mass spectrometry on a Reflex III MALDI-TOF (Bruker,

- 54 -

Bremen Germany). The peptides were used without further purification.

Enzyme linked immune assay (ELISA).

Biotin-labeled peptides (at the N-terminus) were coated on Streptavidin ELISA plates (EXICON) at 10 µg/ml concentration according to the manufacturer's instructions. Highly specific Horse Radish Peroxidase (HRP)-conjugated anti-human IgG secondary antibodies (Southern Biotech) were used according to the manufacturers' recommendations (dilution: 1,000x). Sera were tested at two serum dilutions, 200X and 1,000X. Following manual coating, peptide plates were processed and analyzed by the Gemini 160 ELISA robot (TECAN) with a built-in ELISA reader (GENIOS, TECAN).

Approximately 110 patients and 60 healthy adult sera were included in the analysis. Following the bioinformatic analysis of selected clones, corresponding peptides were designed and synthesized. In case of epitopes with more than 26 amino acid residues, overlapping peptides were made. All peptides were synthesized with a N-terminal biotin-tag and used as coating reagents on Streptavidin-coated ELISA plates.

The analysis was performed in two steps. First, peptides were selected based on their reactivity with the individual sera, which were included in the serum pools used for preparations of IgG and IgA screening reagents for bacterial surface display. A summary for serum reactivity of 224 peptides representing *S. pneumoniae* epitopes from the genomic screen analysed with 20 human sera (representing 4 different pools of five sera) used for the antigen identification is shown in Table 2. The peptides were compared by the score calculated for each peptide based on the number of positive sera and the extent of reactivity. Peptides range from highly and widely reactive to weakly positive ones. Among the most reactive ones there are known antigens, some of them are also protective in animal challenge models for nasopharyngeal carriage or sepsis (e.g. PspA/SP0117, serine protease/SP0641, histidine triad protein/SP1175). Peptides not displaying a positive reaction were not included in further, more detailed studies.

Second, a large number of not pre-selected individual sera from patients with invasive pneumococcal disease or from healthy adults and children were tested against the peptides showing specific and high reactivity with the screening sera. Seroconversion during disease was tested for highly positive peptides by using three serial serum samples collected longitudinally from patients with invasive pneumococcal disease, the first before disease occurred (pre), the second in the acute phase (within 5 days after onset) and the third in the convalescent phase (> 3 weeks after onset) of the disease. Two representative ELISA experiments are shown with two different patients, displaying seroconversion to multiple peptides, suggesting that epitope-specific antibody levels were low before disease occurred, and were induced in the acute and convalescent phase (Fig. 5). The antigens showing this antibody profile are especially valuable for vaccine development (e.g. SP2216, SP2109, SP1175, SP0117, SP0082).

Example 5: Gene distribution studies with highly immunogenic proteins identified from *S. pneumoniae*.

Experimental procedures

Gene distribution of pneumococcal antigens by PCR. An ideal vaccine antigen would be an antigen that is present in all, or the vast majority of strains of the target organism to which the vaccine is directed. In order to establish whether the genes encoding the identified *Streptococcus pneumoniae* antigens occur ubiquitously in *S. pneumoniae* strains, PCR was performed on a series of independent *S. pneumoniae* isolates with primers specific for the gene of interest. *S. pneumoniae* isolates were obtained covering the serotypes most frequently present in patients as shown in Figure 4A. Oligonucleotide sequences as primers were designed for all identified ORFs yielding products of approximately 1,000 bp, if possible

- 55 -

covering all identified immunogenic epitopes. Genomic DNA of all *S. pneumoniae* strains was prepared as described under Example 2. PCR was performed in a reaction volume of 25 µl using Taq polymerase (1U), 200 nM dNTPs, 10 pMol of each oligonucleotide and the kit according to the manufacturers instructions (Invitrogen, The Netherlands). As standard, 30 cycles (1x: 5min. 95°C, 30x: 30sec. 95°C, 30sec. 56°C, 30sec. 72°C, 1x 4min. 72°C) were performed, unless conditions had to be adapted for individual primer pairs.

Results

All identified genes encoding immunogenic proteins were tested by PCR for their presence in 50 different strains of *S. pneumoniae* (Figure 4A). As an example, figure 4B shows the PCR reaction for SP1604 with all indicated 50 strains. As clearly visible, the gene is present in all strains analysed. The PCR fragment from a type 14 strain was sequenced and showed that of 414 bp, 6 bp are different as compared to the *S. pneumoniae* type 4 strain, resulting in three amino acid difference between the two isolates.

From a total of 50 genes analysed, 31 were present in all strains tested, while 9 genes were absent in more than 10 of the tested 50 strains (Table 3). Several genes (SP0667, SP0930) showed variation in size and were not present in all strain isolates. Some genes showed variation in size, but were otherwise conserved in all tested strains. Sequencing of the generated PCR fragment from one strain and subsequent comparison to the type 4 strain confirmed the amplification of the correct DNA fragment and revealed a degree of sequence divergence as indicated in Table 3. Importantly, many of the identified antigens are well conserved in all strains in sequence and size and are therefore novel vaccine candidates to prevent infections by pneumococci.

Example 6: Characterization of immune sera obtained from mice immunized with highly immunogenic proteins/peptides from *S. pneumoniae* displayed on the surface of *E. coli.*

Experimental procedures

Generation of immune sera from mice
E. coli clones harboring plasmids encoding the platform protein fused to a *S. pneumoniae* peptide, were grown in LB medium supplemented with 50µg/ml Kanamycin at 37°C. Overnight cultures were diluted 1:10, grown until an OD_{600} of 0.5 and induced with 0.2 mM IPTG for 2 hours. Pelleted bacterial cells were suspended in PBS buffer and disrupted by sonication on ice, generating a crude cell extract. According to the OD_{600} measurement, an aliquot corresponding to $5x10^7$ cells was injected into NMRI mice i.v., followed by a boost after 2 weeks. Serum was taken 1 week after the second injection. Epitope specific antibody levels were measured by peptide ELISA.

In vitro expression of antigens
Expression of antigens by *in vitro* grown *S. pneumoniae* serotype 4 was tested by immunoblotting. Different growth media and culture conditions were tested to detect the presence of antigens in total lysates and bacterial culture supernatants. Expression was considered confirmed when a specific band corresponding to the predicted molecular weight and electrophoretic mobility was detected.

Cell surface staining
Flow cytometric analysis was carried out as follows. Bacteria were grown under culture conditions, which resulted in expression of the antigen as shown by the immunoblot analysis. Cells were washed twice in Hanks Balanced Salt Solution (HBSS) and the cell density was adjusted to approximately 1×10^6 CFU in 100µl HBSS, 0.5% BSA. After incubation for 30 to 60 min at 4°C with mouse antisera diluted 50 to 100-fold, unbound antibodies were washed away by centrifugation in excess HBSS, 0.5% BSA. Secondary goat anti-mouse antibody (F(ab')₂ fragment specific) labeled with fluorescein (FITC) was incubated with the cells at 4°C for 30 to 60 min. After washing, cells were fixed with 2% paraformaldehyde. Bound antibodies were detected using a Becton Dickinson FACScan flow cytometer and data further analyzed

- 56 -

with the computer program CELLQuest. Negative control sera included mouse pre-immune serum and mouse polyclonal serum generated with lysates prepared from IPTG induced E. coli cells transformed with plasmids encoding the genes lamB or fhuA without S. pneumoniae genomic insert.

Bactericidal (killing) assay

Murine macrophage cells (RAW246.7 or P388.D1) and bacteria were incubated and the loss of viable bacteria after 60 min was determined by colony counting. In brief, bacteria were washed twice in Hanks Balanced Salt Solution (HBSS) and the cell density was adjusted to approximately $1X\ 10^5$ CFU in 50μl HBSS. Bacteria were incubated with mouse sera (up to 25%) and guinea pig complement (up to 5%) in a total volume of 100μl for 60min at 4°C. Pre-opsonized bacteria were mixed with macrophages (murine cell line RAW264.7 or P388.D1; $2X\ 10^6$ cells per 100μl) at a 1:20 ratio and were incubated at 37°C on a rotating shaker at 500 rpm. An aliquot of each sample was diluted in sterile water and incubated for 5 min at room temperature to lyse macrophages. Serial dilutions were then plated onto Todd-Hewitt Broth agar plates. The plates were incubated overnight at 37°C, and the colonies were counted with the Countermat flash colony counter (IUL Instruments). Control sera included mouse pre-immune serum and mouse polyclonal serum generated with lysates prepared from IPTG induced E. coli transformed with plasmids harboring the genes lamB or fhuA without S. pneumoniae genomic insert.

Results

In vitro expression of antigens. The expression of the antigenic proteins was analyzed *in vitro* in S. pneumoniae serotype 4 by using sera raised against E. coli clones harboring plasmids encoding the platform protein fused to a S. pneumoniae peptide. First, the presence of specific antibodies was determined by peptide ELISA and/or immunoblotting using the E. coli clone expressing the given epitope embedded in LamB or FhuA platform proteins. Positive sera were then analysed by immunblotting using total bacterial lysates and culture supernatants prepared from S. pneumoniae serotype 4 strain (data not shown). This analysis served as a first step to determine whether a protein is expressed at all, and if, under which growth conditions, in order to evaluate surface expression of the polypeptide by FACS analysis. It was anticipated based on literature data that not all proteins would be expressed under *in vitro* conditions.

Cell surface staining of S. pneumoniae. Cell surface accessibility for several antigenic proteins was subsequently demonstrated by an assay based on flow cytometry. Streptococci were incubated with preimmune and polyclonal mouse sera raised against S. pneumoniae lysate or E. coli clones harboring plasmids encoding the platform protein fused to a S. pneumoniae peptide, follow by detection with fluorescently tagged secondary antibody. As shown in Fig. 6A, antisera raised against S. pneumoniae lysate contains antibodies against surface components, demonstrated by a significant shift in fluorescence of the S. pneumoniae serotype 4 cell population. Similar cell surface staining of S. pneumoniae sertype 4 cells was observed with polyclonal sera raised against peptides of many of the pneumococcal antigens identified (Fig. 6B and Table 4.). In some instancies, a subpopulation of the bacteria was not stained, as indicated by the detection of two peaks in the histograms (Fig. 6B). This phenomenon may be a result of differential expression of the gene products during the growth of the bacterium, insufficient antibody levels or partial inhibition of antibody binding caused by other surface molecules or plasma proteins.

In vitro bactericidal activity. Opsonophagocytic killing is the cornerstone of host defense against extracellular bacteria, such as S. pneumoniae. Cell surface binding of antibodies to bacterial antigens are opsonizing and induce killing (bactericidal) by phagocytic cells (macrophages and neutrophil granulocytes) if the antibodies induced by the particular antigens can bind activated complement components (C3bi). It has been shown that anti-pneumococcal bactericidal activity of human sera measured in *in vitro* assays can be correlated with *in vivo* protection of vaccinated individuals (Romero-Steiner, S. et al., 1999). In Figure 7 examples are shown and in Table 4 a summary is presented on bactericidal activity measured by antigen-specific antibodies generated in mice with corresponding

epitopes. According to these data, several of the novel pneumococcal antigens induce functional antibodies (e.g. SP0082, SP2216, SP2136, SP0454, SP0069, SP0369, etc.). Importantly, a well-known protective pneumoniae antigen, PspA (SP0117) is proved to be strongly positive in the very same assay.

These experiments confirmed the bioinformatic prediction that many of the proteins are exported due to their signal peptide sequence and in addition showed that they are present on the cell surface of *S. pneumoniae* serotype 4. They also confirm that these proteins are available for recognition by human antibodies with functional properties and make them valuable candidates for the development of a vaccine against pneumococcal diseases.

Example 7: Identification of pneumococcal antigens inducing protective immune responses

Experimental procedures

Expression of recombinant pneumococcal proteins
Cloning of genes/DNA fragments: The gene/DNA fragment of interest was amplified from the genomic DNA of S pneumoniae (strain T4, Capsular type 4) by PCR using gene specific primers. Apart from the gene specific part, the primers had restriction sites that aided in a directional cloning of the amplified PCR product. The gene annealing (specific) part of the primer ranged between 15-24 bases in length. The PCR products obtained were digested with the appropriate restriction enzyme and cloned into pET28b(+) vector (NOVAGEN). Once the recombinant plasmid was confirmed to contain the gene of interest, E coli BL21 star® cells (INVITROGEN) that served as expression hosts were transformed. These cells were optimized to efficiently express the gene of interest. Expression and purification of proteins: E coli BL21 star® cells harbouring the recombinant plasmid was grown until log phase in a required culture volume. Once the OD600nm of 0.8 was reached the culture was induced with 1 mM IPTG for 3 hours at 37°C. The cells were harvested by centrifugation, lysed by a combination of freeze-thaw method followed by disruption of cells with 'Bug-buster®, NOVAGEN'. The lysate was separated by centrifugation into soluble (supernatant) and insoluble (pellet) fractions. Depending on the location of the protein different purification strategies was followed. In case the protein was in the soluble fraction, purification of the protein was done by binding the above supernatant with Ni-Agarose beads (Ni-NTA-Agarose®, QIAGEN). Due to the presence of the penta Histidine (HIS) at the C or N or both termini of the expressed protein, it bound to the Ni-agarose while the other contaminating proteins were washed from the column by wash buffer. The proteins were eluted by 100 mM immidazole and the eluate was concentrated, assayed by Bradford for protein concentration and checked by PAGE and Western blot. In case the protein was present in the insoluble fraction the pellet was solubilized in buffer containing 8 M Urea. The purification was done under denaturing conditions (in buffer containing 8M Urea) using the same materials and procedure as mentioned above. The eluate was concentrated and dialyzed to remove all the urea in a gradual stepwise manner. The proteins were checked by SDS-PAGE and concentrations measured by Bradford method.

Animal protection studies
Animals: C3H (HeNHsd; A, B, C, D: agouti, wild type, inbred) female mice were used. Active immunization: 50 µg of recombinant proteins were injected subcutaneously and adjuvanted with Complete Freud Adjuvant (CFA). Animals were boosted twice with the same amount of protein, but adjuvanted with Incomplete Freund Adjuvant (IFA) at days 14 and 28. A well-known protective antigen PspA (SP0117) was used as a positive control, while nonimmunized (PBS or CFA/IFA adjuvant injected) mice served as negative controls. Antibody titers were measured at days 35-38 by ELISA using the respective recombinant proteins, and were determined to be in the range of 200.00-1.000.000 (end-point titer). Passive immunization: Näive mice were injected with 150-300µl mouse sera intraperitoneally 2hrs before intraperitoneal inoculation with *S. pneumoniae.* Bacterial challenge: A frozen glycerol stock of *S. pneumoniae* serotype 6B was prepared and used for all experiments. The approximate estimated cell number was determined by OD600nm measurements. In order to determine the real viable cell numbers

- 58 -

present in the inoculum prepared from the frozen glycerol stock cultures, cfus were determined via plating using six serial dilutions. 10^4-10^8 bacteria/mouse was injected either intravenously in the tail vein or intraperitoneally. Protective effect of immunizations was measured by monitoring survival rates for 2 to 3 weeks post-challenge and was expressed in % of total number of animals (10/ group).

Results

In the present invention six different pneumococcal antigens identified by bacterial surface display were determined to have protective effect in a mouse sepsis/lethality model. The best levels of protection were achieved by immunization with recombinant antigens representing the SP2190, SP2216 and SP0667 protein, while SP0368, SP1732 and SP0498 displayed lower levels of protection (Fig. 8). The protective effect was mediated by antibodies as it has been demonstrated by passive serum transfer experiments (Fig. 9). Näive mice receiving specific anti-SP2190 and anti-SP2216 antibodies were protected from death relative to mice from the negative control groups. Importantly, the combination of these antigens resulted in improved protection, as it has been shown in Fig. 9. Passive immunization with 150 µl immune serum generated either with recombinant SP2190 or with recombinant SP2216 (supplemented with 150µl näive serum) resulted in lower survival rate compared to serum therapy with 100µl each of specific antisera (supplemented with 100µl näive serum). These experiments strongly support that combination of these antigens has beneficial effects in vaccination against pneumococcal diseases.

Since the antigens used for immunization were derived from a serotype 4 strain and the challenge strain was a serotype 6B, these experiments established that the antigens were cross-protective.

The SP2216, SP2190 and SP1732 recombinant proteins detected the highest levels of antibodies in sera of patients convalescing from invasive pneumococcal diseases, as well as in those of healthy individuals exposed to Pneumococcus (children in the household) (data not shown). The most frequently identified antigen in bacterial surface display screens was the SP2216 protein. It was a special interest to compare the protectivity of the subdomains of this protein selected (N-terminal amino acid sequences) or not selected (C-terminal amino acid sequences) by human antibodies (Fig. 10A). Upon immunization with the two different domains (expressed as recombinant antigens) it became evident that the immunogenic part of the SP2216 protein carried the protective potential, while the non-selected domain was ineffective and comparable to the negative control (Fig. 10B). Based on this experiment the epitopes detected by bacterial surface display identifies protective epitopes and regions of bacterial proteins and this information can be used for rational design of subunit vaccines based on the antigens described in the present invention.

Example 8: Determination of sequence conservation of protective antigens

Experimental procedures

Immunoblotting
Total bacterial lysate and culture supernatant samples were prepared from *in vitro* grown *S. pneumoniae* strains. 60 (clinical isolates) representing 48 different serotypes were included in the study. Approximately 25µg total protein/lane was separated by SDS-PAGE using the BioRad Mini-Protean 3 Cell electrophoresis system and proteins transferred to nitrocellulose membrane (ECL, Amersham Pharmacia). After overnight blocking in 5% milk, hyperimmune mouse sera generated by immunization with the recombinant proteins SP2216, SP1732 and SP2190 (and SP0117/PspA as internal control) derived from serotype 4 strain were added at 5,000x dilution, and HRPO labeled anti-mouse IgG was used for detection.

DNA sequencing

The genes of SP1732, SP2190 and SP22126 were amplified from the genomic DNA of S. pneumoniae (serotype4, 6B, 9V, 14, 18C, 19F and23F) by PCR using a proofreading polymerase Expand (ROCHE). Gene specific primers, ranging between 27-31 bases in length, were used to amplify the entire open reading frames. The PCR products obtained were cloned into pCR®2.1-TOPO vector (Invitrogen). The recombinant plasmid DNA was purified using a QIAprep® miniprep kit (Qiagen) before the sequence was confirmed (MWG). In addition to the seven serotypes, the gene of SP2216 from other 41 different serotypes was amplified by PCR and the purified PCR products were sequenced.

Results

Identification of conserved antigens inducing antibodies that are cross-reactive with different clinical isolates is crucial for the development effective vaccines. It is especially relevant for protein-based vaccines targeting pneumococcal diseases, since more then 90 different serotypes of *Streptococcus pnuemoniae (Pneumococcus)* have been associated with human infections.

In a thorough analysis it was determined that the antibodies induced by SP2216, SP2190 and SP1732 all derived from a serotype 4 strain broadly cross-reacted with all the different serotypes tested in the immunoblot analysis (Fig. 11). Notably, the SP2190 antigen that showed variation in electrophoretic mobility (indicating different sizes) preserved the antibody reactivity strongly suggesting that immunodominant epitopes are conserved. In contrast, we detected lower cross-reactivity with anti-PspA antibodies that is in accordance with the known differences in immunogenic amino acid sequences of this antigen.

In order to directly address the question whether the identified protective antigens are conserved among the different serotypes of *S. pneumoniae*, DNA sequence analysis was performed on the SP2216, SP1732 and SP2190 genes. SP2216 and SP1732 are highly conserved, only few amino acid changes were detected. The SP2216 gene was sequenced from 47 different clinical isolates representing 47 different *S. pneumoniae* serotypes and only single amino acid exchanges were detected and only in 2 of the analysed strains (Fig. 12). In the SP1732 gene one or two amino acid exchanges were detected in the majority (in four of the six) of strains analysed (Fig. 13). The 2190 antigen showed a great variability in the amino acid sequences of the corresponding genes as it is shown in Fig. 14. The insertions and deletions makes it difficult to calculate an exact amino acid homology among the different SP2190 variants, but it can be estimated to be between approx. 60 and 90%. However, the amino acid identity was sufficient to induce cross-reactive and cross-protective antibodies based on the experiments presented in Fig. 8,9 and 11.

References

Adamou, J., et al. (2001). Infect Immun 69: 949-58.

Altschul, S., et al. (1990). Journal of Molecular Biology 215: 403-10.

Bennett, D., et al. (1995). J Mol Recognit 8: 52-8.

Brown, J., et al. (2001). Infect Immun 69: 6702-6.

Burnie, J., et al. (1998). J Antimicrob Chemother 41: 319-22.

Clackson, T., et al. (1991). Nature 352: 624-8.

Devereux, J., et al. (1984). Nucleic acids research 12: 387-95.

Di Guilmi, A., et al. (2002). EMBO Rep 3: 728-34.

Doherty, E., et al. (2001). Annu Rev Biophys Biomol Struct 30: 457-475.

Eisenbraun, M., et al. (1993). DNA Cell Biol 12: 791-7.

Epidemiology and Prevention of Vaccine-Preventable Diseases, 7 th Edition-Second Printing (The Pink Book). The Pink Book is published by the Centers for Disease Control and Prevention, U.S. Department of Health and Human Services.

Etz, H., et al. (2001). J Bacteriol 183: 6924-35.

Ganz, T. (1999). Science 286: 420-421.

Georgiou, G. (1997). Nature Biotechnology 15: 29-34.

Gray, B., et al. (1979). J Infect Dis 140: 979-83.

Gray, B., et al. (1986). Pediatr Infect Dis 5: 201-7.

Hashemzadeh-Bonehi, L., et al. (1998). Mol Microbiol 30: 676-678.

Hausdorff, W., et al. (2001). Lancet 357: 950-2.

Heinje, von G. (1987) e.g. Sequence Analysis in Molecular Biology, Acedimic Press

Hemmer, B., et al. (1999). Nat Med 5: 1375-82.

Hoe, N., et al. (2001). J Infect Dis 183: 633-9.

Hornef, M., et al. (2002). Nat Immunol 3: 1033-40.

Hoskins, J., et al. (2001). J Bacteriol 183: 5709-17.

Hyde, T., et al. (2001). JAMA 286: 1857-62.

Jedrzejas, M. (2001). Microbiol Mol Biol Rev 65: 187-207.

Johanson, K., et al. (1995). J Biol Chem 270: 9459-71.

Jones, P., et al. (1986). Nature 321: 522-5.

Kajava, A., et al. (2000). J Bacteriol 182: 2163-9.

Kohler, G., et al. (1975). Nature 256: 495-7.

Kolaskar, A., et al. (1990). FEBS Lett 276: 172-4.

Lewin, A., et al. (2001). Trends Mol Med 7: 221-8.

Marks, J., et al. (1992). Biotechnology (N Y) 10: 779-83.

McCafferty, J., et al. (1990). Nature 348: 552-4.

McCormick, A., et al. (2003). Nat Med 9: 424-30.

McDaniel, L., et al. (1991). Infect Immun 59: 222-8.

Navarre, W., et al. (1999). Microbiol Mol Biol Rev 63: 174-229.

Okano, H., et al. (1991). J Neurochem 56: 560-7.

Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression; CRC Press, Boca Tation, FL (1988) for a description of these molecules

Orange, M., et al. (1993). Pediatr Infect Dis J 12: 244-6.

Pelton, S., et al. (2003). Vaccine 21: 1562-71.

Phillips-Quagliata, J., et al. (2000). J Immunol 165: 2544-55.

Rammensee, H., et al. (1999). Immunogenetics 50: 213-9.

Roche, H., et al. (2003). Infect Immun 71: 1033-41.

Romero-Steiner, S., et al. (1999). Clin Infect Dis 29: 281-8.

Rosenow, C., et al. (1997). Mol Microbiol 25: 819-29.

Seeger, C., et al. (1984). Proc Natl Acad Sci U S A 81: 5849-52.

Shibuya, A., et al. (2000). Nature Immunology 1: 441-6.

Skerra, A. (1994). Gene 151: 131-5.
Talkington, D., et al. (1996). Microb Pathog 21: 17-22.
Tang, D., et al. (1992). Nature 356: 152-4.
Tempest, P., et al. (1991). Biotechnology (N Y) 9: 266-71.
Tettelin, H., et al. (2001). Science 293: 498-506.
Tourdot, S., et al. (2000). Eur J Immunol 30: 3411-21.
Whitney, C., et al. (2000). N Engl J Med 343: 1917-24.
Wiley, J., et al. (1987) Current Protocols in Molecular Biology.
Wizemann, T., et al. (2001). Infect Immun 69: 1593-8.

Table 1: Immunogenic proteins identified by bacterial surface display.

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
SP0008	hypothetical protein	4-11,35-64,66-76,101-108,111-119	G:15	57-114	1, 145
SP0032	DNA polymerase I (polA)	5-27,32-64,92-102,107-113,119-125,133-139,148-162,177-187,195-201,207-214,241-251,254-269,285-300,302-309,317-324,332-357,365-404,411-425,443-463,470-477,479-487,506-512,515-520,532-547,556-596,603-610,616-622,624-629,636-642,646-665,667-674,687-692,708-720,734-739,752-757,798-820,824-851,856-865	H:39, I:6, L:2	732-763	2, 146
SP0069	Choline binding protein I	14-21,36-44,49-66,102-127,162-167,177-196	G:1, H:2, I:1, K:44, L:3, M:1	45-109 145-172	3, 147
SP0071	immunoglobulin A1 protease (iga-1)	17-35,64-75,81-92,100-119,125-172,174-183,214-222,230-236,273-282,287-303,310-315,331-340,392-398,412-420,480-505,515-523,525-546,553-575,592-598,603-609,617-625,631-639,644-651,658-670,681-687,691-704,709-716,731-736,739-744,750-763,774-780,784-791,799-805,809-822,859-870,880-885,907-916,924-941,943-949,973-986,1010-1016,1026-1036,1045-1054,1057-1062,1082-1088,1095-1102,1109-1120,1127-1134,1140-1146,1152-1159,1169-1179,1187-1196,1243-1251,1262-1273,1279-1292,1306-1312,1332-1343,1348-1364,1379-1390,1412-1420,1427-1436,1458-1468,1483-1503,1524-1549,1574-1588,1614-1619,1672-1685,1697-1707,1711-1720,1738-1753,1781-1787,1796-1801,1826-1843	A:3, C:1, D:9, E:9, F:4, G:21, I:34, K:61, L:20, M:2	132-478 508-592 1753-1810	4, 148
SP0082	Cell wall surface anchor	15-43,49-55,71-77,104-110,123-130,162-171,180-192,199-205,219-227,246-254,264-270,279-287,293-308,312-322,330-342,349-356,369-377,384-394,401-406,416-422,432-439,450-460,464-474,482-494,501-508,521-529,536-546,553-558,568-574,584-591,602-612,616-626,634-646,653-660,673-681,688-698,705-710,720-726,736-749,833-848	C:9, E:4, F:2, I:26, L:4, M:67	1-199 200-337 418-494 549-647	5, 149
SP0107	LysM domain protein	9-30,65-96,99-123,170-178	A :3, B :16, C :15, D :1, E :5, F :178,	1-128	6, 150

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
			M :1		
SP0117	pneumococcal surface protein A (pspA)	7-32,34-41,96-106,127-136,154-163,188-199,207-238,272-279,306-312,318-325,341-347,353-360,387-393,399-406,434-440,452-503,575-580,589-601,615-620,635-640,654-660,674-680,696-701,710-731	A:13, B:11, C:10, D:4, E:31, F:6, G:33, H:13, I:9, K:64, L:32, M:46	1-548 660-691	7, 151
SP0191	hypothetical protein	4-19,35-44,48-59,77-87,93-99,106-111,130-138,146-161	E:1, I:2	78-84	8, 152
SP0197	dihydrofolate synthetase, putative	24-30,36-43,64-86,93-99,106-130,132-145,148-165,171-177,189-220,230-249,251-263,293-300,302-312,323-329,338-356,369-379,390-412	L :9	179-193	9, 153
SP0212	Ribosomal protein L2	30-39,61-67,74-81,90-120,123-145,154-167,169-179,182-197,200-206,238-244,267-272	L:10	230-265	10, 154
SP0222	Ribosomal protein S14	14-20,49-65,77-86	H:14, L:8, M:3	2-68	11, 155
SP0239	Conserved hypothetical protein	4-9,26-35,42-48,53-61,63-85,90-101,105-111,113-121,129-137,140-150,179-188,199-226,228-237,248-255,259-285,299-308,314-331,337-343,353-364,410-421,436-442	L:2, M:1	110-144	12, 156
SP0251	formate acetyltransferase, putative	36-47,55-63,94-108,129-134,144-158,173-187,196-206,209-238,251-266,270-285,290-295,300-306,333-344,346-354,366-397,404-410,422-435,439-453,466-473,515-523,529-543,554-569,571-585,590-596,607-618,627-643,690-696,704-714,720-728,741-749,752-767,780-799	G:2, H:7, I:1, M:5	225-247 480-507	13, 157
SP0295	ribosomal protein S9 (rpsI)	16-25,36-70,80-93,100-106	I:4	78-130	14, 158
SP0330	sugar binding transcriptional regulator RegR	18-27,41-46,50-57,65-71,79-85,93-98,113-128,144-155,166-178,181-188,201-207,242-262,265-273,281-295,303-309,318-327	G:1, H:1, L:4	36-64	15, 159
SP0368	cell wall surface anchor family protein	7-29,31-44,50-59,91-96,146-153,194-201,207-212,232-238,264-278,284-290,296-302,326-353,360-370,378-384,400-405,409-418,420-435,442-460,499-506,529-534,556-562,564-576,644-651,677-684,687-698,736-743,759-766,778-784,808-814,852-858,874-896,920-925,929-935,957-965,1003-1012,1021-1027,1030-1044,1081-1087,1101-1111,1116-1124,1148-1159,1188-	D:1, H:3, I:1, L:1, M:3	1-70 154-189 922-941 1445-1462 1483-1496	16, 160

S. *pneumoniae* antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		1196,1235-1251,1288-1303,1313-1319,1328-1335,1367-1373,1431-1437,1451-1458,1479-1503,1514-1521,1530-1540,1545-1552,1561-1568,1598-1605,1617-1647,1658-1665,1670-1676,1679-1689,1698-1704,1707-1713,1732-1738,1744-1764			
SP0369	Penicillin binding protein 1A	6-51,81-91,104-113,126-137,150-159,164-174,197-209,215-224,229-235,256-269,276-282,307-313,317-348,351-357,376-397,418-437,454-464,485-490,498-509,547-555,574-586,602-619	B:1, E:1, L:13, M:1	452-530	17, 161
SP0374	hypothetical protein	25-31,39-47,49-56,99-114,121-127,159-186,228-240,253-269,271-279,303-315,365-382,395-405,414-425,438-453	E:4, I:1, L:3	289-384	18, 162
SP0377	Choline binding protein C	9-24,41-47,49-54,68-78,108-114,117-122,132-140,164-169,179-186,193-199,206-213,244-251,267-274,289-294,309-314,327-333	G:5, H:4, I:1, K:88, L:3, M:8	209-249 286-336	19, 163
SP0378	choline binding protein J (cbpJ)	9-28,53-67,69-82,87-93,109-117,172-177,201-207,220-227,242-247,262-268,305-318,320-325	K:47, L:6, M:5	286-306	20, 164
SP0390	choline binding protein G (cbpG)	4-10,26-39,47-58,63-73,86-96,98-108,115-123,137-143,148-155,160-176,184-189,194-204,235-240,254-259,272-278	G:1, K:69, M:6	199-283	21, 165
SP0454	hypothetical protein	4-26,33-39,47-53,59-65,76-83,91-97,104-112,118-137,155-160,167-174,198-207,242-268,273-279,292-315,320-332,345-354,358-367,377-394,403-410,424-439,445-451,453-497,511-518,535-570,573-589,592-601,604-610	H:1, I:1, L:6	202-242	22, 166
SP0463	cell wall surface anchor family protein	8-30,36-45,64-71,76-82,97-103,105-112,134-151,161-183,211-234,253-268,270-276,278-284,297-305,309-315,357-362,366-372,375-384,401-407,409-416,441-455,463-470,475-480,490-497,501-513,524-537,552-559,565-576,581-590,592-600,619-625,636-644,646-656	A:1, B:2, C:4, E:1, F:4,	316-419	23, 167
SP0466	sortase, putative	4-17,52-58,84-99,102-110,114-120,124-135,143-158,160-173,177-196,201-216,223-250,259-267,269-275	B:1, M:2	1-67	24, 168
SP0468	Sortase, putative	6-46,57-67,69-80,82-133,137-143,147-168,182-187,203-209,214-229,233-242,246-280	G:24, H:20, L:1	53-93	25, 169
SP0498	endo-beta-N-acetylglucosaminidase, putative	7-40,50-56,81-89,117-123,202-209,213-218,223-229,248-261,264-276,281-288,303-308,313-324,326-332,340-346,353-372,434-443,465-474,514-523,556-564,605-616,620-626,631-636,667-683,685-699,710-719,726-732,751-756,760-771,779-788,815-828,855-867,869-	B:5, C:1, E:2, F:1, G:2	1226-1309 1455-1536 1538-1605	26, 170

- 65 -

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		879,897-902,917-924,926-931,936-942,981-1000,1006-1015,1017-1028,1030-1039,1046-1054,1060-1066,1083-1092,1099-1112,1122-1130,1132-1140,1148-1158,1161-1171,1174-1181,1209-1230,1236-1244,1248-1254,1256-1267,1269-1276,1294-1299,1316-1328,1332-1354,1359-1372,1374-1380,1384-1390,1395-1408,1419-1425,1434-1446,1453-1460,1465-1471,1474-1493,1505-1515,1523-1537,1547-1555,1560-1567,1577-1605,1633-1651			
SP0509	type I restriction-modification system, M subunit	4-10,31-39,81-88,106-112,122-135,152-158,177-184,191-197,221-227,230-246,249-255,303-311,317-326,337-344,346-362,365-371,430-437,439-446,453-462,474-484	I:2	449-467	27, 171
SP0519	dnaJ protein (dnaJ)	9-15,24-35,47-55,122-128,160-177,188-196,202-208,216-228,250-261,272-303,318-324,327-339,346-352,355-361,368-373	A:1, D:2, H:2	108-218 344-376	28, 172
SP0529	BlpC ABC transporter (blpB)	6-14,17-48,55-63,71-90,99-109,116-124,181-189,212-223,232-268,270-294,297-304,319-325,340-348,351-370,372-378,388-394,406-415,421-434	A:1, B:3, C:3, D:1, F:4,	177-277	29, 173
SP0564	hypothetical protein	21-39,42-61,65-75,79-85,108-115	H:3	11-38	30, 174
SP0609	amino acid ABC transporter, amino acid-binding protein	4-17,26-39,61-76,103-113,115-122,136-142,158-192,197-203,208-214,225-230,237-251	I:3	207-225	31, 175
SP0613	metallo-beta-lactamase superfamily protein	5-11,27-36,42-53,62-70,74-93,95-104,114-119,127-150,153-159,173-179,184-193,199-206,222-241,248-253,257-280,289-295,313-319,322-342,349-365,368-389,393-406,408-413,426-438,447-461,463-470,476-495,532-537,543-550	I:12	225-246	32, 176
SP0641	Serine protease	4-29,68-82,123-130,141-147,149-157,178-191,203-215,269-277,300-307,327-335,359-370,374-380,382-388,393-400,410-417,434-442,483-492,497-503,505-513,533-540,564-569,601-607,639-647,655-666,693-706,712-718,726-736,752-758,763-771,774-780,786-799,806-812,820-828,852-863,884-892,901-909,925-932,943-948,990-996,1030-1036,1051-1059,1062-1068,1079-1086,1105-1113,1152-1162,1168-1179,1183-1191,1204-1210,1234-1244,1286-1295,1318-1326,1396-1401,1451-1460,1465-1474,1477-1483,1488-1494,1505-1510,1514-1521,1552-1565,1593-1614,1664-1672,1677-1685,1701-1711,1734-1745,1758-1770,1784-1798,1840-	A:19, B:72, C:34, D:5, E:21, F:86, G:26, H:86, I:17, L:130, M:29	1-348 373-490 573-767 903-1043 1155-1198 1243-1482 1550-1595 1682-1719 1793-1921 2008-2110	33, 177

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		1847,1852-1873,1885-1891,1906-1911,1931-1939,1957-1970,1977-1992,2014-2020,2026-2032,2116-2134			
SP0648	beta-galactosidase (bgaA)	10-35,39-52,107-112,181-188,226-236,238-253,258-268,275-284,296-310,326-338,345-368,380-389,391-408,410-418,420-429,444-456,489-505,573-588,616-623,637-643,726-739,741-767,785-791,793-803,830-847,867-881,886-922,949-956,961-980,988-1004,1009-1018,1027-1042,1051-1069,1076-1089,1108-1115,1123-1135,1140-1151,1164-1179,1182-1191,1210-1221,1223-1234,1242-1250,1255-1267,1281-1292,1301-1307,1315-1340,1348-1355,1366-1373,1381-1413,1417-1428,1437-1444,1453-1463,1478-1484,1490-1496,1498-1503,1520-1536,1538-1546,1548-1570,1593-1603,1612-1625,1635-1649,1654-1660,1670-1687,1693-1700,1705-1711,1718-1726,1729-1763,1790-1813,1871-1881,1893-1900,1907-1935,1962-1970,1992-2000,2006-2013,2033-2039,2045-2051,2055-2067,2070-2095,2097-2110,2115-2121,2150-2171,2174-2180,2197-2202,2206-2228	C:1, E:1, F:1, G:1, H:4, I:1, M:2	1526-1560	34, 178
SP0664	Zinc metalloprotease ZmpB, putative	4-17,35-48,54-76,78-107,109-115,118-127,134-140,145-156,169-174,217-226,232-240,256-262,267-273,316-328,340-346,353-360,402-409,416-439,448-456,506-531,540-546,570-578,586-593,595-600,623-632,662-667,674-681,689-705,713-724,730-740,757-763,773-778,783-796,829-835,861-871,888-899,907-939,941-955,957-969,986-1000,1022-1028,1036-1044,1068-1084,1095-1102,1118-1124,1140-1146,1148-1154,1168-1181,1185-1190,1197-1207,1218-1226,1250-1270,1272-1281,1284-1296,1312-1319,1351-1358,1383-1409,1422-1428,1438-1447,1449-1461,1482-1489,1504-1510,1518-1527,1529-1537,1544-1551,1569-1575,1622-1628,1631-1637,1682-1689,1711-1718,1733-1740,1772-1783,1818-1834,1859-1872	A:9, B:25, C:13, D:7, E:14, F:77, G:12, H:10, K:67, L:13, M:6	1-64 128-495	35, 179
SP0667	pneumococcal surface protein, putative	8-28,32-37,62-69,119-125,137-149,159-164,173-189,200-205,221-229,240-245,258-265,268-276,287-293,296-302,323-329	A:72, B:80, C:90, D:20, E:12, F:53	1-95	36, 180
SP0688	UDP-N-acetylmuramoylalanine--D-glutamate ligase	9-18,25-38,49-63,65-72,74-81,94-117,131-137,139-146,149-158,162-188,191-207,217-225,237-252,255-269,281-293,301-326,332-342,347-354,363-370,373-	L:3	75-107	37, 181

S. *pneumoniae* antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		380,391-400,415-424,441-447			
SP0749	branched-chain amino acid ABC transporter	4-24,64-71,81-87,96-116,121-128,130-139,148-155,166-173,176-184,203-215,231-238,243-248,256-261,280-286,288-306,314-329	E:2, I:8, L:8	67-148	38, 182
SP0770	ABC transporter, ATP-binding protein	4-10,19-37,46-52,62-81,83-89,115-120,134-139,141-151,168-186,197-205,209-234,241-252,322-335,339-345,363-379,385-393,403-431,434-442,447-454,459-465,479-484,487-496	L:2	404-420	39, 183
SP0785	conserved hypothetical protein	10-35,46-66,71-77,84-93,96-122,138-148,154-172,182-213,221-233,245-263,269-275,295-301,303-309,311-320,324-336,340-348,351-359,375-381	C:1, E:2, I:1	111-198	40, 184
SP0914	nodulin-related protein, truncation	14-25,30-42,47-61,67-75,81-91,98-106,114-122,124-135,148-193,209-227	L:2	198-213	41, 185
SP0930	choline binding protein E (cbpE)	5-18,45-50,82-90,97-114,116-136,153-161,163-171,212-219,221-227,240-249,267-281,311-317,328-337,375-381,390-395,430-436,449-455,484-495,538-543,548-554,556-564,580-586,596-602	E:4, G:2, H:1, I:2, K:5	493-606	42, 186
SP0943	Gid protein (gid)	9-25,28-34,37-44,61-68,75-81,88-96,98-111,119-133,138-150,152-163,168-182,186-194,200-205,216-223,236-245,257-264,279-287,293-304,311-318,325-330,340-346,353-358,365-379,399-409,444-453	E:2, L:24	303-391	43, 187
SP0952	alanine dehydrogenase, authentic frameshift (ald)	16-36,55-61,66-76,78-102,121-130,134-146,150-212,221-239,255-276,289-322,329-357	G:3, H:4	29-59	44, 188
SP1003	conserved hypothetical protein (PAT)	8-27,68-74,77-99,110-116,124-141,171-177,202-217,221-228,259-265,275-290,293-303,309-325,335-343,345-351,365-379,384-394,406-414,423-437,452-465,478-507,525-534,554-560,611-624,628-651,669-682,742-747,767-778,782-792,804-812,820-836	A:2, B:5, C:8, D:5, E:13, F:3, M:2	79-231 359-451	45, 189
SP1004	Conserved hypothetical protein	5-28,39-45,56-62,67-74,77-99,110-117,124-141,168-176,200-230,237-244,268-279,287-299,304-326,329-335,348-362,370-376,379-384,390-406,420-429,466-471,479-489,495-504,529-541,545-553,561-577,598-604,622-630,637-658,672-680,682-688,690-696,698-709,712-719,724-736,738-746,759-769,780-786,796-804,813-818,860-877,895-904,981-997,1000-1014,1021-1029	A:5, B:4, C:4, D:9, E:12, F:4, H:3, I:1, L:1	1-162 206-224 254-350 414-514 864-938	46, 190

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
SP1124	glycogen synthase (glgA)	4-11,19-49,56-66,68-101,109-116,123-145,156-165,177-185,204-221,226-234,242-248,251-256,259-265,282-302,307-330,340-349,355-374,377-383,392-400,422-428,434-442,462-474	M:1	266-322	47, 191
SP1154	IgA1 protease	14-43,45-57,64-74,80-87,106-127,131-142,145-161,173-180,182-188,203-210,213-219,221-243,245-254,304-311,314-320,342-348,354-365,372-378,394-399,407-431,436-448,459-465,470-477,484-490,504-509,531-537,590-596,611-617,642-647,723-734,740-751,754-762,764-774,782-797,807-812,824-831,838-845,877-885,892-898,900-906,924-935,940-946,982-996,1006-1016,1033-1043,1051-1056,1058-1066,1094-1108,1119-1126,1129-1140,1150-1157,1167-1174,1176-1185,1188-1201,1209-1216,1220-1228,1231-1237,1243-1248,1253-1285,1288-1297,1299-1307,1316-1334,1336-1343,1350-1359,1365-1381,1390-1396,1412-1420,1427-1439,1452-1459,1477-1484,1493-1512,1554-1559,1570-1578,1603-1608,1623-1630,1654-1659,1672-1680,1689-1696,1705-1711,1721-1738,1752-1757,1773-1780,1817-1829,1844-1851,1856-1863,1883-1895,1950-1958,1974-1990	A:6, B:2, C:9, D:3, E:4, F:2, G:6, H:4, I:13, L:12	172-354 384-448 464-644 648-728 1357-1370	48, 192
SP1174	conserved domain protein (PAT)	8-27,68-74,77-99,110-116,124-141,169-176,201-216,220-227,258-264,274-289,292-302,308-324,334-342,344-350,364-372,377-387,399-407,416-429,445-458,471-481,483-500,518-527,547-553,604-617,621-644,662-675,767-778,809-816	B:14, C:17, D:6, E:18, F:16, I:1, K:5, L:1, M:8	15-307 350-448 496-620	49, 193
SP1175	conserved domain protein	4-17,24-29,53-59,62-84,109-126,159-164,189-204,208-219,244-249,274-290,292-302,308-324,334-342,344-350,378-389,391-397,401-409,424-432,447-460,470-479,490-504,521-529,538-544,549-555,570-577,583-592,602-608,615-630,635-647,664-677,692-698,722-731,733-751,782-790,793-799	A:1, B:4, C:3, D:3, E:9, F:2, H:2, M:4	56-267 337-426 495-601	50, 194
SP1221	type II restriction endonuclease	12-22,49-59,77-89,111-121,136-148,177-186,207-213,217-225,227-253,259-274,296-302,328-333,343-354,374-383,424-446,448-457,468-480,488-502,507-522,544-550,553-560,564-572,587-596,604-614,619-625,629-635,638-656,662-676,680-692,697-713,720-738,779-786,833-847,861-869,880-895,897-902,911-917,946-951,959-967,984-990,992-1004,1021-1040,1057-	G:2, H:1, K:1, L:4	381-403	51, 195

S. *pneumoniae* antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		1067,1073-1080			
SP1227	DNA-binding response regulator	4-10,26-31,46-56,60-66,70-79,86-94,96-102,109-118,132-152,164-187,193-206,217-224	E:1, L:3	81-149	52, 196
SP1241	amino acid ABC transporter, amino acid-binding pro	4-21,26-37,48-60,71-82,109-117,120-128,130-136,142-147,181-187,203-211,216-223,247-255,257-284,316-325,373-379,395-400,423-435,448-456,479-489,512-576,596-625,641-678,680-688,692-715	B:2, C:1, E:2, I:1	346-453	53, 197
SP1287	signal recognition particle protein (ffh)	10-16,25-31,34-56,58-69,71-89,94-110,133-176,186-193,208-225,240-250,259-266,302-307,335-341,376-383,410-416	B :8, G :8, H :3, M :1	316-407	54, 198
SP1330	N-acetylmannosamine-6-P epimerase, putative (nanE)	11-29,42-56,60-75,82-88,95-110,116-126,132-143,145-160,166-172,184-216	L:45	123-164	55, 199
SP1374	Chorismate sythetase (aroC)	11-29,54-63,110-117,139-152,158-166,172-180,186-193,215-236,240-251,302-323,330-335,340-347,350-366,374-381	G:1, L:29, M:14	252-299	56, 200
SP1378	conserved hypothetical protein	18-27,35-42,50-56,67-74,112-136,141-153,163-171,176-189,205-213,225-234,241-247,253-258,269-281,288-298,306-324,326-334,355-369,380-387	H:2	289-320	57, 201
SP1429	peptidase, U32 family	7-15,19-41,56-72,91-112,114-122,139-147,163-183,196-209,258-280,326-338,357-363,391-403,406-416	H:4	360-378	58, 202
SP1478	oxidoreductase, aldo/keto reductase family	11-18,29-41,43-49,95-108,142-194,204-212,216-242,247-256,264-273	H:11	136-149	59, 203
SP1518	conserved hypothetical protein	18-24,33-40,65-79,89-102,113-119,130-137,155-161,173-179,183-203,205-219,223-231,245-261,267-274,296-306,311-321,330-341,344-363,369-381,401-408,415-427,437-444,453-464,472-478,484-508,517-524,526-532,543-548	A:10, E:4, G:5, H:1	59-180	60, 204
SP1522	conserved domain protein	5-13,52-65,67-73,97-110,112-119,134-155	B:4, C:6, E:1, H:7, L:3	45-177	61, 205
SP1527	oligopeptide ABC transporter	6-28,34-43,57-67,75-81,111-128,132-147,155-163,165-176,184-194,208-216,218-229,239-252,271-278,328-334,363-376,381-388,426-473,481-488,492-498,507-513,536-546,564-582,590-601,607-623	A:1, B:1, C:4, F:1, G:26, H:18, I:10, L:2, M:1	148-269 420-450 610-648	62, 206

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
SP1573	lysozyme (lytC)	4-12,20-38,69-75,83-88,123-128,145-152,154-161,183-188,200-213,245-250,266-272,306-312,332-339,357-369,383-389,395-402,437-453,455-470,497-503	A:40, B:27, C:24, D:2, E:6, G:11, K:1	1-112	63, 207
SP1604	hypothetical protein	35-59,74-86,111-117,122-137	A:1, C:3, E:1, G:1, I:1	70-154	64, 208
SP1661	cell division protein DivIVA	26-42,54-61,65-75,101-107,123-130,137-144,148-156,164-172,177-192,213-221,231-258	E :2	157-249	65, 209
SP1664	ylmF protein (ylmF)	29-38,61-67,77-87,94-100,105-111,118-158	B:1, C:42, I:3	1-97	66, 210
SP1676	N-acetylneuraminate lyase, putative	7-21,30-48,51-58,60-85,94-123,134-156,160-167,169-183,186-191,216-229,237-251,257-267,272-282,287-298	H:2	220-243	67, 211
SP1687	neuraminidase B (nanB)	6-29,34-47,56-65,69-76,83-90,123-134,143-151,158-178,197-203,217-235,243-263,303-309,320-333,338-348,367-373,387-393,407-414,416-427,441-457,473-482,487-499,501-509,514-520,530-535,577-583,590-602,605-612,622-629,641-670,678-690	B:3, E:2, L:1, M:2	37-71 238-307	68, 212
SP1693	neuraminidase A (nanA)	7-40,121-132,148-161,196-202,209-215,221-235,248-255,271-280,288-295,330-339,395-409,414-420,446-451,475-487,556-563,568-575,580-586,588-595,633-638,643-648,652-659,672-685,695-700,710-716,737-742,749-754,761-767,775-781,796-806,823-835,850-863,884-890,892-900,902-915,934-941	C:3, D:5, E:3, F:1, G:7, H:1, I:3, K:20, L:4	406-521	69, 213
SP1732	serine/threonine protein kinase	9-18,24-46,51-58,67-77,85-108,114-126,129-137,139-146,152-165,173-182,188-195,197-204,217-250,260-274,296-313,343-366,368-384,427-434,437-446,449-455,478-484,492-506,522-527,562-591,599-606,609-618,625-631,645-652	E:2, H:1	577-654	70, 214
SP1735	methionyl-tRNA formyltransferase (fmt)	13-20,26-37,41-53,56-65,81-100,102-114,118-127,163-188,196-202,231-238,245-252,266-285,293-298,301-306	K:13, M:13	19-78	71, 215
SP1759	preprotein translocase, SecA subunit (secA-2)	10-23,32-42,54-66,73-91,106-113,118-127,139-152,164-173,198-207,210-245,284-300,313-318,330-337,339-346,354-361,387-393,404-426,429-439,441-453,467-473,479-485,496-509,536-544,551-558,560-566,569-574,578-588,610-615,627-635,649-675,679-690,698-716,722-734,743-754,769-780,782-787	I:6, L:2, M:2	480-550	72, 216

S. *pneumoniae* antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
SP1772	cell wall surface anchor family protein	6-39,42-50,60-68,76-83,114-129,147-162,170-189,197-205,217-231,239-248,299-305,338-344,352-357,371-377,380-451,459-483,491-499,507-523,537-559,587-613,625-681,689-729,737-781,785-809,817-865,873-881,889-939,951-975,983-1027,1031-1055,1063-1071,1079-1099,1103-1127,1151-1185,1197-1261,1269-1309,1317-1333,1341-1349,1357-1465,1469-1513,1517-1553,1557-1629,1637-1669,1677-1701,1709-1725,1733-1795,1823-1849,1861-1925,1933-1973,1981-2025,2029-2053,2061-2109,2117-2125,2133-2183,2195-2219,2227-2271,2275-2299,2307-2315,2323-2343,2347-2371,2395-2429,2441-2529,2537-2569,2577-2601,2609-2625,2633-2695,2699-2737,2765-2791,2803-2867,2889-2913,2921-2937,2945-2969,2977-2985,2993-3009,3023-3045,3073-3099,3111-3167,3175-3215,3223-3267,3271-3295,3303-3351,3359-3367,3375-3425,3437-3461,3469-3513,3517-3541,3549-3557,3565-3585,3589-3613,3637-3671,3683-3747,3755-3795,3803-3819,3827-3835,3843-3951,3955-3999,4003-4039,4043-4115,4123-4143,4147-4171,4195-4229,4241-4305,4313-4353,4361-4377,4385-4393,4401-4509,4513-4557,4561-4597,4601-4718,4749-4768	B:9, C:1, D:1, F:13, G:1, H:3, I:1, L:1, M:2	74-171 452-559 2951-3061	73, 217
SP1804	general stress protein 24, putative	16-22,30-51,70-111,117-130,137-150,171-178,180-188,191-196	I:4	148-181	74, 218
SP1888	oligopeptide ABC transporter, ATP-binding protein AmiE	6-19,21-46,50-56,80-86,118-126,167-186,189-205,211-242,244-267,273-286,290-297,307-316,320-341	H:1	34-60	75, 219
SP1891	oligopeptide ABC transporter,	5-26,33-43,48-54,58-63,78-83,113-120,122-128,143-152,157-175,185-192,211-225,227-234,244-256,270-281,284-290,304-310,330-337,348-355,362-379,384-394,429-445,450-474,483-490,511-520,537-546,548-554,561-586,590-604,613-629	A:2, B:3, E:1, F:1, G:13, H:8	149-186 285-431 573-659	76, 220
SP1937	Autolysin (lytA)	5-26,49-59,61-67,83-91,102-111,145-157,185-192,267-272,279-286,292-298,306-312	D:3, F:1, G:1, H:2, K:11, M:1	134-220 235-251 254-280	77, 221
SP1954	serine protease, subtilase family, authentic frame	5-19,72-79,83-92,119-124,140-145,160-165,167-182,224-232,240-252,259-270,301-310,313-322,332-343,347-367,384-398,416-429,431-446,454-461	C:43, E:6, I:4, K:21, L:50	1-169	78, 222
SP1980	cmp-binding-factor 1	8-17,26-31,56-62,75-83,93-103,125-131,135-141,150-	H:9	127-168	79, 223

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
	(cbf1)	194,205-217,233-258,262-268,281-286			
SP1992	cell wall surface anchor family protein	6-12,69-75,108-115,139-159,176-182,194-214	B:5, C:1, F:4, I:1	46-161	80, 224
SP1999	catabolite control protein A (ccpA)	6-13,18-27,39-48,51-59,66-73,79-85,95-101,109-116,118-124,144-164,166-177,183-193,197-204,215-223,227-236,242-249,252-259,261-270,289-301,318-325	I:2	12-58	81, 225
SP2021	glycosyl hydrolase	4-10,26-32,48-60,97-105,117-132,138-163,169-185,192-214,219-231,249-261,264-270,292-308,343-356,385-392,398-404,408-417,435-441	L:3	24-50	82, 226
SP2027	Conserved hypothetical protein	10-40,42-48,51-61,119-126	A:1, E:1, G:19, H:12, I:16, L:5	1-118	83, 227
SP2039	conserved hypothetical protein	5-17,40-58,71-83,103-111,123-140,167-177,188-204	G:1, L:3	116-128	84, 228
SP2048	Conserved hypothetical protein	4-9,11-50,57-70,112-123,127-138	I:1, L:4	64-107	85, 229
SP2051	Conpetence protein CglC	9-39,51-67	D:1, G:3, I:8, L:26	1-101	86, 230
SP2092	UTP-glucose-1-phosphate uridylyltransferase (galU)	5-14,17-25,28-46,52-59,85-93,99-104,111-120,122-131,140-148,158-179,187-197,204-225,271-283,285-293	H:2	139-155	87, 231
SP2099	Penicillin binding protein 1B	42-70,73-90,92-108,112-127,152-164,166-172,181-199,201-210,219-228,247-274,295-302,322-334,336-346,353-358,396-414,419-425,432-438,462-471,518-523,531-536,561-567,576-589,594-612,620-631,665-671,697-710,718-731,736-756,765-771,784-801	A:1, B:9, C:11, D:1, E:6, F:1, H:4, K:1	626-653	88, 232
SP2108	Maltose ABC transporter	8-28,41-51,53-62,68-74,79-85,94-100,102-108,114-120,130-154,156-162,175-180,198-204,206-213,281-294,308-318,321-339,362-368,381-386,393-399,407-415	G:10, H:1, L:10, M:1	2-13	89, 233
SP2120	hypothetical protein	4-39,48-65,93-98,106-112,116-129	I:2	10-36	90, 234
SP2128	transketolase, N-terminal subunit	25-32,35-50,66-71,75-86,90-96,123-136,141-151,160-179,190-196,209-215,222-228,235-242,257-263,270-280	H:2	209-247	91, 235
SP2136	choline binding protein PcpA	5-29,31-38,50-57,62-75,83-110,115-132,168-195,197-206,216-242,249-258,262-269,333-340,342-350,363-368,376-392,400-406,410-421,423-430,436-442,448-454,460-466,471-476,491-496,511-516,531-536,551-	C:3, F:1, G:24, H:32, I:13, K:177,	27-70 219-293 441-504 512-584	92, 236

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
		556,571-576,585-591,599-605	L:34, M:18		
SP2141	glycosyl hydrolase-related protein	4-12,14-34,47-75,83-104,107-115,133-140,148-185,187-196,207-212,224-256,258-265,281-287,289-296,298-308,325-333,345-355,365-371,382-395,424-435,441-457,465-472,483-491,493-505,528-534,536-546,552-558,575-584,589-600,616-623	L:3	576-591	93, 237
SP2180	conserved hypothetical protein	4-76, 78-89, 91-126, 142-148, 151-191, 195-208, 211-223, 226-240, 256-277, 279-285, 290-314, 317-323, 358-377, 381-387, 391-396, 398-411, 415-434, 436-446, 454-484, 494-512, 516-523, 538-552, 559-566, 571-577, 579-596, 599-615, 620-627, 635-644, 694-707, 720-734, 737-759, 761-771	L:3	313-329	94, 238
SP2190	choline binding protein A (cbpA)	7-38,44-49,79-89,99-108,117-123,125-132,137-146,178-187,207-237,245-255,322-337,365-387,398-408,445-462,603-608,623-628,644-650,657-671,673-679	A:6, B:12, C:9, D:6, E:30, F:8, G:65, H:72, I:76, K:222, L:99, M:37	111-566	95, 239
SP2194	ATP-dependent Clp protease, ATP-binding subunit	6-20,22-35,39-45,58-64,77-117,137-144,158-163,205-210,218-224,229-236,239-251,263-277,299-307,323-334,353-384,388-396,399-438,443-448,458-463,467-478,481-495,503-509,511-526,559-576,595-600,612-645,711-721,723-738,744-758,778-807	H:1	686-720	96, 240
SP2201	choline binding protein D (cbpD)	10-33,35-41,72-84,129-138,158-163,203-226,243-252,258-264,279-302,322-329,381-386,401-406,414-435	B:4, C:3, D:1, E:7, F:1, G:1, H:2, K:26, M:1	184-385	97, 241
SP2204	ribosomal protein L9	4-9,19-24,41-47,75-85,105-110,113-146	H:3, L:4	45-62	98, 242
SP2216	secreted 45 kd protein - homology to glucan binding protein (GbpB) S.mutant	4-25,52-67,117-124,131-146,173-180,182-191,195-206,215-221,229-236,245-252,258-279,286-291,293-302,314-320,327-336,341-353,355-361,383-389	A:130, B:414, C:450, D:162, E:166, F:284, G:90, H:16, I:4, K:10, L:29,	1-285	99, 243

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
			M:11		
SP-NRF1	Choline binding protein	14-32,38-50,73-84,93-105,109-114	H:1	40-70	100, 244
ARF0408	Hypothetical protein	5-26	L:3	22-34	101, 245
ARF0441	Hypothetical protein	23-28	H:3	13-39	102, 246
ARF0690	Hypothetical protein	8-14	L:2	21-34	103, 247
ARF0878	Hypothetical protein	4-13,20-29,44-50,59-74	H:3	41-69	104, 248
ARF0921	Hypothetical protein	4-9, 19-42, 48-59, 71-83	M:4	57-91	105, 249
ARF1153	Hypothetical protein	4-14	M:7	10-28	106, 250
ARF1515	Hypothetical protein	22-28,32-42,63-71,81-111,149-156,158-167,172-180,182-203,219-229	G:4, H:5	27-49	107, 251
ARF1519	Hypothetical protein	17-27	H:3	23-32	108, 252
ARF1905	Hypothetical protein	18-24	H:2	28-38	109, 253
ARF2044	Hypothetical protein	9-15	G:2, H:5	13-27	110, 254
ARF2155	Hypothetical protein	13-22	H:3	18-29	111, 255
ARF2199	Hypothetical protein	17-26	M:3	2-11	112, 256
CRF0129	Hypothetical protein	4-33	L:4	16-32	113, 257
CRF0200	Hypothetical protein	4-10,37-43,54-84,92-127	H:5, L:1	15-62	114, 258
CRF0236	Hypothetical protein	4-14,20-32,35-60,69-75,79-99,101-109,116-140	L:3	124-136	115, 259
CRF0394	Hypothetical protein	none	H:7	2-13	116, 260
CRF0408	Hypothetical protein	4-13,28-42	L:11	42-57	117, 261
CRF0430	Hypothetical protein	4-14,27-44	G:4, H:8	14-35	118, 262
CRF0498	Hypothetical protein	4-12	H:4	1-27	119, 263
CRF0519	Hypothetical protein	4-18,39-45,47-74	G:5, H:3	35-66	120, 264
CRF0573	Hypothetical protein	8-20,43-77	I:3, L:9	17-36	121, 265
CRF0713	Hypothetical protein	4-30,35-45,51-57	L:3	35-49	122, 266
CRF0722	Hypothetical protein	4-24,49-57	G:18	15-34	123, 267
CRF0764	Hypothetical protein	4-22	L:4	8-27	124, 268
CRF1079	Hypothetical protein	13-25,32-59,66-80	H:5	21-55	125, 269
CRF1248	Hypothetical protein	4-10,24-33,35-42,54-65,72-82,98-108	H:1	15-30	126, 270
CRF1398	Hypothetical protein	8-19	H:1, L:3	17-47	127, 271
CRF1412	Hypothetical protein	12-18,40-46	L:8	31-52	128, 272
CRF1467	Hypothetical protein	4-20,35-78,83-102,109-122	I:4	74-86	129, 273
CRF1484	Hypothetical protein	7-17,21-41,46-63	I:5	2-20	130, 274
CRF1587	Hypothetical protein	30-37	G:3, H:3,	2-33	131, 275

S. pneumoniae antigenic protein	Putative function (by homology)	predicted immunogenic aa**	No. of selected clones per ORF and screen	Location of identified immunogenic region (aa)	Seq. ID (DNA, Prot.)
			L:4		
CRF1606	Hypothetical protein	4-13,17-25	L:3	1-15	132, 276
CRF1623	Hypothetical protein	17-31,44-51	M:6	20-51	133, 277
CRF1625	Hypothetical protein	20-30	L:10	5-23	134, 278
CRF1640	Hypothetical protein	13-33,48-71	I:5	92-110	135, 279
CRF1702	Hypothetical protein	4-9,50-69,76-88,96-106,113-118	L:6	12-34	136, 280
CRF1825	Hypothetical protein	4-24	L:11	6-26	137, 281
CRF1883	Hypothetical protein	7-26	H:61, L:77	14-30	138, 282
CRF1991	Hypothetical protein	9-39,46-68,75-82,84-103	H:6, L:2	26-44	139, 283
CRF1992	Hypothetical protein	4-30,33-107	M:7	58-84	140, 284
CRF2004	Hypothetical protein	4-12	L:3	9-51	141, 285
CRF2030	Hypothetical protein	12-18,29-37	H:5, L:1, M:1	6-37	142, 286
CRF2065	Hypothetical protein	4-21,33-52,64-71	I:1, M:6	16-37	143, 287
CRF2232	Hypothetical protein	9-19	L:3	2-30	144, 288

Table 2. Immunogenicity of epitopes in peptide ELISA

Peptides	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	S	from	to	Seq ID
ARF0408.1	nd					nd		nd																15	20	37	245
ARF0441.1	nd					nd		nd																26	8	27	246
ARF0680.1	nd					nd		nd																20	10	27	247
ARF0878.1						nd					nd	nd												22	42	59	248
ARF0878.2						nd					nd	nd												11	52	69	248
ARF0921.1						nd					nd	nd												15	63	80	249
ARF0921.2						nd					nd	nd												14	74	91	249
ARF1153.1	nd					nd		nd																19	11	28	250
ARF1515.1	nd					nd		nd																16	28	49	251
ARF1519.1	nd					nd		nd																15	15	32	252
ARF1905.1	nd					nd		nd																21	4	20	253
ARF2044.1	nd					nd		nd																18	10	27	254
ARF2155.1	nd					nd		nd																15	17	34	255
ARF2199.1	nd					nd		nd																19	1	18	256
CRF0129.1	nd					nd		nd																15	16	33	257
CRF0200.1	nd					nd		nd																12	16	36	258
CRF0200.2	nd					nd		nd																9	30	49	258
CRF0200.3	nd					nd		nd																10	43	62	258
CRF0236.1						nd					nd	nd												19	122	139	259
CRF0394.1	nd					nd		nd																20	1	18	260
CRF0408.1	nd					nd		nd																19	41	58	261
CRF0430.1	nd					nd		nd																15	15	35	262
CRF0498.1	nd					nd		nd																21	2	27	263
CRF0573.1						nd					nd	nd												22	18	36	265
CRF0713.1	nd					nd		nd																20	34	51	266
CRF0764.1	nd					nd		nd																16	9	27	268
CRF1079.1	nd					nd		nd																27	22	47.	269
CRF1398.1	nd					nd		nd																24	18	36	271
CRF1398.2	nd					nd		nd																21	29	47	271
CRF1412.1						nd					nd	nd												9	32	52	272
CRF1467.1	nd					nd		nd																	72	89	273
CRF1484.1	nd					nd		nd																	3	20	274
CRF1587.1	nd					nd		nd																23	3	21	275
CRF1587.2	nd					nd		nd																21	15	33	275
CRF1606.1	nd					nd		nd																22	1	18	276
CRF1625.1	nd					nd		nd																23	6	23	278
CRF1640.1	nd					nd		nd																18	93	110	279
CRF1702.1						nd					nd	nd												18	13	34	280
CRF1825.1	nd					nd		nd																24	7	26	281
CRF1825.2	nd					nd		nd																24	9	26	281
CRF1883.1	nd					nd		nd																20	16	33	282
CRF1991.1	nd					nd		nd																24	27	44	283
CRF1992.1	nd							nd		nd														19	67	84	284
CRF2004.1	nd					nd		nd																20	10	33	285
CRF2004.2	nd					nd		nd																22	26	50	285
CRF2030.1	nd					nd		nd																21	7	25	286

ID							
CRF2030.2		20	19	37	286
CRF2065.1		21	17	37	287
CRF2232.1		23	3	20	288
CRF2232.2		19	13	30	288
SP0008.1		28	62	80	145
SP0008.2		21	75	93	145
SP0069.1		17	92	108	147
SP0071.1		15	332	349	148
SP0071.2		13	177	200	148
SP0071.3		4	1755	1777	148
SP0082.1		4	109	133	149
SP0082.2		24	149	174	149
SP0082.3		26	260	285	149
SP0082.4		20	460	485	149
SP0107.1		10	26	47	150
SP0107.2		11	42	64	150
SP0117.1		18	22	41	151
SP0117.2		57	35	54	151
SP0117.3		64	115	130	151
SP0117.4		17	306	325	151
SP0117.5		15	401	420	151
SP0117.6		68	454	478	151
SP0222.1		15	22	45	155
SP0368.1		22	156	174	160
SP0368.2		32	924	940	160
SP0368.2b		12	1485	1496	160
SP0368.3		14	1447	1462	160
SP0368.4		11	1483	1498	160
SP0369.1		16	457	475	161
SP0377.1		11	302	325	163
SP0378.1		17	288	305	164
SP0390.1		11	244	266	165
SP0390.2		24	260	282	165
SP0454.1		64	204	225	166
SP0454.2		38	220	241	166
SP0463.1		20	324	345	167
SP0463.2		26	340	361	167
SP0463.3		19	356	377	167
SP0463.4		9	372	393	167
SP0463.5		13	388	408	167
SP0466.1		22	39	64	168
SP0468.1		21	54	76	169
SP0468.2		24	70	92	169
SP0498.1		13	1227	1247	170
SP0498.2		8	1539	1559	170
SP0498.3		16	1554	1574	170
SP0498.4		13	1569	1589	170
SP0498.5		15	1584	1604	170
SP0498.6		11	1242	1262	170
SP0498.8		24	1272	1292	170

ID				
SP0498.9	9	1287	1308	170
SP0498.10	63	1456	1477	170
SP0498.11	59	1472	1494	170
SP0498.12	20	1488	1510	170
SP0498.13	10	1505	1526	170
SP0519.1	25	351	368	172
SP0529.1	6	179	200	173
SP0529.2	23	195	216	173
SP0529.3	7	211	232	173
SP0529.4	8	227	248	173
SP0529.5	18	243	263	173
SP0564.1	63	13	37	174
SP0609.1	16	208	224	175
SP0641.1	34	42	64	177
SP0641.2	61	59	81	177
SP0641.3	0	304	328	177
SP0641.4	36	323	348	177
SP0641.5	76	465	489	177
SP0641.6	64	968	992	177
SP0641.7	52	1399	1418	177
SP0641.8	25	1412	1431	177
SP0641.9	18	2092	2111	177
SP0648.1	11	1528	1547	178
SP0648.2	24	1541	1560	178
SP0664.1	28	184	200	179
SP0664.2	18	367	388	179
SP0664.3	15	382	403	179
SP0664.4	14	409	429	179
SP0664.5	20	425	444	179
SP0664.6	17	438	457	179
SP0667.1	22	27	50	180
SP0667.2	16	45	67	180
SP0749.2	24	114	131	182
SP0770.1	17	405	419	183
SP0785.1	11	113	134	184
SP0785.2	10	129	150	184
SP0785.3	8	145	166	184
SP0785.4	11	161	182	184
SP0785.5	8	177	198	184
SP0930.1	30	495	515	186
SP0943.1	19	346	358	187
SP1004.1	21	208	224	190
SP1154.1	33	178	194	192
SP1154.2	28	202	223	192
SP1154.3	21	217	238	192
SP1154.4	19	288	308	192
SP1154.5	16	1355	1372	192
SP1175.1	56	57	78	194
SP1241.1	26	347	369	197
SP1241.2	18	364	386	197

SP1241.3	25	381	403	197
SP1241.4	21	398	420	197
SP1241.5	26	415	437	197
SP1241.6	16	432	452	197
SP1287.1	20	347	372	198
SP1330.1	18	147	163	199
SP1374.1	15	263	288	200
SP1429.1	20	361	377	202
SP1518.1	19	82	104	204
SP1518.2	28	99	121	204
SP1518.3	21	116	138	204
SP1518.4	26	133	155	204
SP1518.5	19	150	171	204
SP1522.1	12	110	130	205
SP1522.2	13	125	145	205
SP1527.1	22	613	631	206
SP1527.2	20	626	644	206
SP1527.3	16	196	213	206
SP1604.1	13	78	100	208
SP1604.2	20	95	117	208
SP1604.3	15	112	134	208
SP1604.4	25	129	151	208
SP1661.1	18	158	180	209
SP1661.2	10	175	197	209
SP1661.3	13	192	214	209
SP1661.4	16	209	231	209
SP1661.5	12	226	248	209
SP1664.1	24	30	50	210
SP1664.2	26	45	65	210
SP1664.3	11	60	79	210
SP1693.1	20	431	455	213
SP1693.2	18	450	474	213
SP1732.1	25	579	601	214
SP1732.2	15	596	618	214
SP1732.3	19	613	635	214
SP1732.4	22	630	653	214
SP1772.1	27	920	927	217
SP1772.1b	12	98	119	217
SP1772.2	20	114	135	217
SP1772.3	11	130	151	217
SP1772.4	30	146	167	217
SP1772.5	19	162	182	217
SP1888.1	16	36	59	219
SP1891.1	14	194	216	220
SP1891.2	12	381	404	220
SP1937.1	20	236	251	221
SP1937.2	26	255	279	221
SP1954.1	15	80	100	222
SP1954.2	14	141	164	222
SP1980.1	18	128	154	223

Label		20	82	100	224
SP1992.1	nd	20	82	100	224
SP1992.2	nd	23	95	116	224
SP1992.3	nd	20	111	134	224
SP2027.1	nd	12	55	76	227
SP2027.2	nd	16	71	92	227
SP2027.3	nd	21	87	110	227
SP2048.1		21	91	106	229
SP2051.1		25	74	96	230
SP2092.1	nd	22	140	157	231
SP2108.1	nd	13	4	13	233
SP2136.1	nd	5	41	65	236
SP2136.2	nd	32	499	523	236
SP2190.1	nd	13	122	146	239
SP2190.2	nd	29	191	215	239
SP2190.3	nd	26	288	313	239
SP2190.4	nd	36	445	469	239
SP2190.5	nd	24	511	535	239
SP2201.1	nd	34	347	368	241
SP2204.1	nd	16	46	61	242
SP2216.1		16	15	37	243
SP2216.2		28	32	57	243
SP2216.3		14	101	121	243
SP2216.4		31	115	135	243
SP2216.5		22	138	158	243
SP2216.6		12	152	172	243
SP2216.7		46	220	242	243
SP2216.8		12	236	258	243

Table 3: Gene distribution in *S. pneumoniae* strains.

ORF	Common name	Gene distribution (present of 50)	Amino acid substitutions (in serotype 14 strain)*	Homology ()	Seq. ID (DNA, Prot.)
SP0008	hypothetical protein	n.d.	n.d.		1, 145
SP0032	DNA polymerase I (polA)	n.d.	n.d.		2, 146
SP0069	Choline binding protein I	7	0/166#		3, 147
SP0071	immunoglobulin A1 protease (iga-1)	7	0/477#		4, 148
SP0082	Cell wall surface anchor	50	5/385		5, 149
SP0107	LysM domain protein	50	1/173		6, 150
SP0117	pneumococcal surface protein A (pspA)	n.d.	n.d.		7, 151
SP0191	hypothetical protein	n.d.	n.d.		8, 152
SP0197	dihydrofolate synthetase, putative	n.d.	n.d.		9, 153
SP0212	Ribosomal protein L2	50	0/232		10, 154
SP0222	Ribosomal protein S14	n.d.	n.d.		11, 155
SP0239	Conserved hypothetical protein	n.d.	n.d.		12, 156
SP0251	formate acetyltransferase, putative	n.d.	n.d.		13, 157
SP0295	ribosomal protein S9 (rpsI)	50	1/121		14, 158
SP0330	sugar binding transcriptional regulator RegR	n.d.	n.d.		15, 159
SP0368	cell wall surface anchor family protein	46	4/422#		16, 160
SP0369	Penicillin binding protein 1A	50	1/346		17, 161
SP0374	hypothetical protein	n.d.	n.d.		18, 162
SP0377	Choline binding protein C	29	0/114		19, 163
SP0378	choline binding protein J (cbpJ)	50	2/104		20, 164
SP0390	choline binding protein G (cbpG)	50	3/171#		21, 165
SP0454	hypothetical protein	48	1/303#		22, 166
SP0463	cell wall surface anchor family protein	10	0/298#		23, 167
SP0466	sortase, putative	44	4/243#		24, 168
SP0468	Sortase, putative	18	0/254#		25, 169
SP0498	endo-beta-N-acetylglucosaminidase, putative	50	4/334		26, 170
SP0509	type I restriction-modification system, M subunit	n.d.	n.d.		27, 171
SP0519	dnaJ protein (dnaJ)	50	2/312		28, 172
SP0529	BlpC ABC transporter (blpB)	50	6/306		29, 173
SP0564	hypothetical protein	50	1/127		30, 174

ORF	Common name	Gene distribution (present of 50)	Amino acid substitutions (in serotype 14 strain)*	Homology 0	Seq. ID (DNA, Prot.)
SP0609	amino acid ABC transporter, amino acid-binding pro	50	0/232		31, 175
SP0613	metallo-beta-lactamase superfamily protein	n.d.	n.d.		32, 176
SP0641	Serine protease	n.d.	n.d.		33, 177
SP0648	beta-galactosidase (bgaA)	50	0/304		34, 178
SP0664	Zinc metalloprotease ZmpB, putative	n.d.	n.d.		35, 179
SP0667	pneumococcal surface protein, putative	45	18/297		36, 180
SP0688	UDP-N-acetylmuramoylalanine--D-glutamate ligase	n.d.	n.d.		37, 181
SP0749	branched-chain amino acid ABC transporter	50	4/303		38, 182
SP0770	ABC transporter, ATP-binding protein	50	0/307		39, 183
SP0785	conserved hypothetical protein	50	0/304		40, 184
SP0914	nodulin-related protein, truncation	n.d.	n.d.		41, 185
SP0930	choline binding protein E (cbpE)	47	17/294		42, 186
SP0943	Gid protein (gid)	n.d.	n.d.		43, 187
SP0952	alanine dehydrogenase, authentic frameshift (ald)	n.d.	n.d.		44, 188
SP1003	conserved hypothetical protein (PAT)	n.d.	n.d.		45, 189
SP1004	Conserved hypothetical protein	n.d.	n.d.		46, 190
SP1124	glycogen synthase (glgA)	n.d.	n.d.		47, 191
SP1154	IgA1 protease	28	13/470; 80missing		48, 192
SP1174	conserved domain protein (PAT)	n.d.	n.d.		49, 193
SP1175	conserved domain protein	n.d.	n.d.		50, 194
SP1221	type II restriction endonuclease	n.d.	n.d.		51, 195
SP1227	DNA-binding response regulator	n.d.	n.d.		52, 196
SP1241	amino acid ABC transporter, amino acid-binding protein	50	0/285		53, 197
SP1287	signal recognition particle protein (ffh)	49	0/300		54, 198
SP1330	N-acetylmannosamine-6-P epimerase, putative (nanE)	14	0/211#		55, 199
SP1374	Chorismate sythetase (aroC)	50	0/289		56, 200
SP1378	conserved hypothetical protein	n.d.	n.d.		57, 201
SP1429	peptidase, U32 family	50	8/305		58, 202
SP1478	oxidoreductase, aldo/keto reductase	n.d.	n.d.		59, 203

ORF	Common name	Gene distribution (present of 50)	Amino acid substitutions (in serotype 14 strain)*	Homology ()	Seq. ID (DNA, Prot.)
	family				
SP1518	conserved hypothetical protein	50	4/313; 3 additional		60, 204
SP1522	conserved domain protein	n.d.	n.d.		61, 205
SP1527	oligopeptide ABC transporter	50	0/463		62, 206
SP1573	lysozyme (lytC)	n.d.	n.d.		63, 207
SP1604	hypothetical protein	50	3/138		64, 208
SP1661	cell division protein DivIVA	50	3/236		65, 209
SP1664	ylmF protein (ylmF)	50	0/164		66, 210
SP1676	N-acetylneuraminate lyase, putative	n.d.	n.d.		67, 211
SP1687	neuraminidase B (nanB)	n.d.	n.d.		68, 212
SP1693	neuraminidase A (nanA)	n.d.	n.d.		69, 213
SP1732	serine/threonine protein kinase	49	2/293		70, 214
SP1735	methionyl-tRNA formyltransferase (fmt)	n.d.	n.d.		71, 215
SP1759	preprotein translocase, SecA subunit (secA-2)	n.d.	n.d.		72, 216
SP1772	cell wall surface anchor family protein	23	12/253#		73, 217
SP1804	general stress protein 24, putative	n.d.	n.d.		74, 218
SP1888	oligopeptide ABC transporter, ATP-binding protein AmiE	n.d.	n.d.		75, 219
SP1891	oligopeptide ABC transporter,	n.d.	n.d.		76, 220
SP1937	Autolysin (lytA)	50	0/275		77, 221
SP1954	serine protease, subtilase family, authentic frame	12	0/305#		78, 222
SP1980	cmp-binding-factor 1 (cbf1)	n.d.	n.d.		79, 223
SP1992	cell wall surface anchor family protein	50	4/197		80, 224
SP1999	catabolite control protein A (ccpA)	n.d.	n.d.		81, 225
SP2021	glycosyl hydrolase	n.d.	n.d.		82, 226
SP2027	Conserved hypothetical protein	n.d.	n.d.		83, 227
SP2039	conserved hypothetical protein	n.d.	n.d.		84, 228
SP2048	Conserved hypothetical protein	50	8/134		85, 229
SP2051	Conpetence protein CglC	50	8/92		86, 230
SP2092	UTP-glucose-1-phosphate uridylyltransferase (galU)	n.d.	n.d.		87, 231
SP2099	Penecillin binding protein 1B	n.d.	n.d.		88, 232
SP2108	Maltose ABC transporter	50	1/279		89, 233
SP2120	hypothetical protein	n.d.	n.d.		90, 234

- 84 -

ORF	Common name	Gene distribution (present of 50)	Amino acid substitutions (in serotype 14 strain)*	Homology ()	Seq. ID (DNA, Prot.)
SP2128	transketolase, N-terminal subunit	n.d.	n.d.		91, 235
SP2136	choline binding protein PcpA	45	1/382		92, 236
SP2141	glycosyl hydrolase-related protein	n.d.	n.d.		93, 237
SP2180	conserved hypothetical protein	n.d.	n.d.		94, 238
SP2190	choline binding protein A (cbpA)	47	for: 48,8%; rev: 2 /17#		95, 239
SP2194	ATP-dependent Clp protease, ATP-binding subunit	50	1/262		96, 240
SP2201	choline binding protein D (cbpD)	50	7/384		97, 241
SP2204	ribosomal protein L9	n.d.	n.d.		98, 242
SP2216	secreted 45 kd protein - homology to glucan binding protein (GbpB) S.mutant	50	0/347		99, 243
SP-NRF1	Choline binding protein	n.d.	n.d.		100, 244
ARF0408	Hypothetical protein	n.d.	n.d.		101, 245
ARF0441	Hypothetical protein	n.d.	n.d.		102, 246
ARF0690	Hypothetical protein	n.d.	n.d.		103, 247
ARF0878	Hypothetical protein	n.d.	n.d.		104, 248
ARF0921	Hypothetical protein	n.d.	n.d.		105, 249
ARF1153	Hypothetical protein	n.d.	n.d.		106, 250
ARF1515	Hypothetical protein	n.d.	n.d.		107, 251
ARF1519	Hypothetical protein	n.d.	n.d.		108, 252
ARF1905	Hypothetical protein	n.d.	n.d.		109, 253
ARF2044	Hypothetical protein	n.d.	n.d.		110, 254
ARF2155	Hypothetical protein	n.d.	n.d.		111, 255
ARF2199	Hypothetical protein	n.d.	n.d.		112, 256
CRF0129	Hypothetical protein	n.d.	n.d.		113, 257
CRF0200	Hypothetical protein	n.d.	n.d.		114, 258
CRF0236	Hypothetical protein	n.d.	n.d.		115, 259
CRF0394	Hypothetical protein	n.d.	n.d.		116, 260
CRF0408	Hypothetical protein	n.d.	n.d.		117, 261
CRF0430	Hypothetical protein	n.d.	n.d.		118, 262
CRF0498	Hypothetical protein	n.d.	n.d.		119, 263
CRF0519	Hypothetical protein	n.d.	n.d.		120, 264
CRF0573	Hypothetical protein	n.d.	n.d.		121, 265
CRF0713	Hypothetical protein	n.d.	n.d.		122, 266
CRF0722	Hypothetical protein	n.d.	n.d.		123, 267
CRF0764	Hypothetical protein	n.d.	n.d.		124, 268

- 85 -

ORF	Common name	Gene distribution (present of 50)	Amino acid substitutions (in serotype 14 strain)*	Homology ()	Seq. ID (DNA, Prot.)
CRF1079	Hypothetical protein	n.d.	n.d.		125, 269
CRF1248	Hypothetical protein	n.d.	n.d.		126, 270
CRF1398	Hypothetical protein	n.d.	n.d.		127, 271
CRF1412	Hypothetical protein	n.d.	n.d.		128, 272
CRF1467.1	Hypothetical protein	n.d.	n.d.		129, 273
CRF1484	Hypothetical protein	n.d.	n.d.		130, 274
CRF1587	Hypothetical protein	n.d.	n.d.		131, 275
CRF1606	Hypothetical protein	n.d.	n.d.		132, 276
CRF1623	Hypothetical protein	n.d.	n.d.		133, 277
CRF1625	Hypothetical protein	n.d.	n.d.		134, 278
CRF1640	Hypothetical protein	n.d.	n.d.		135, 279
CRF1702	Hypothetical protein	n.d.	n.d.		136, 280
CRF1825	Hypothetical protein	n.d.	n.d.		137, 281
CRF1883	Hypothetical protein	n.d.	n.d.		138, 282
CRF1991	Hypothetical protein	n.d.	n.d.		139, 283
CRF1992	Hypothetical protein	n.d.	n.d.		140, 284
CRF2004	Hypothetical protein	n.d.	n.d.		141, 285
CRF2030	Hypothetical protein	n.d.	n.d.		142, 286
CRF2065	Hypothetical protein	n.d.	n.d.		143, 287
CRF2232	Hypothetical protein	n.d.	n.d.		144, 288

Table 4.

ORF	Common Name	FACS	PK
ARF0878	hypothetical protein	+	nd
ARF0921	hypothetical protein	+	nd
CRF0236	hypothetical protein	++	-
CRF0573	hypothetical protein	+	nd
CRF1412	hypothetical protein	+	nd
CRF1702	hypothetical protein	+	nd
CRF1992	hypothetical protein	++	++
SP0008	hypothetical protein	+	-
SP0069	Choline binding protein I	++	++
SP0082	Cell wall surface anchor	+	-
SP0117	pneumococcal surface protein A (pspA)	+++	+++
SP0212	Ribosomal protein L2	+	++
SP0295	ribosomal protein S9 (rpsI)	++	+++
SP0368	cell wall surface anchor family protein	++	+++
SP0369	Penecillin binding protein 1A	++	++
SP0377	Choline binding protein C	++	++
SP0378	choline binding protein J (cbpJ)	++	nd
SP0390	choline binding protein G (cbpG)	++	+
SP0454	hypothetical protein	++	+++
SP0463	cell wall surface anchor family protein	+	++
SP0466	sortase, putative	++	++
SP0468	Sortase, putative	++	++
SP0519	dnaJ protein (dnaJ)	++	+
SP0609	amino acid ABC transporter, amino acid-bind	++	+
SP0641	Serine protease	+	-
SP0664	Zinc metalloprotease ZmpB	+	++
SP0749	branched-chain amino acid ABC transporter	+	+
SP0770	ABC transporter, ATP-binding protein	++	++
SP1154	IgA1 protease	++	++
SP1287	signal recognition particle protein (ffh)	+	++
SP1330	N-acetylmannoseamine-6-P	++	-
SP1429	peptidase, U32 family	+	++
SP1527	oligopeptide ABC transporter	+	++
SP1759	preprotein translocase, SecA subunit (wrong clone!!!)	+	-
SP1772	cell wall surface anchor family protein	+	+
SP1891	oligopeptide ABC transporter	+	++
SP1937	Autolysin (lytA)	+	-
SP1954	serine protease, subtilase family, auth frame	+	++
SP1980	cmp-binding-factor 1 (cbf1)	+	-
SP2108	Maltose ABC transporter	+	++
SP2136	choline binding protein PcpA	+	++
SP2190	choline binding protein A (cbpA)	+	++
SP2194	ATP-dependent Clp protease, ATP-bind subu	++	++
SP2201	choline binding protein D (cbpD)	+	++
SP2216	secreted 45 kd protein	+	++

Claims:

1.   An isolated nucleic acid molecule encoding a hyperimmune serum reactive antigen or a fragment
      thereof comprising a nucleic acid sequence, which is selected from the group consisting of:
      a)   a nucleic acid molecule having at least 70% sequence identity to a nucleic acid molecule selected
           from Seq ID No 1, 101-144.
      b)   a nucleic acid molecule which is complementary to the nucleic acid molecule of a),
      c)   a nucleic acid molecule comprising at least 15 sequential bases of the nucleic acid molecule of a)
           or b)
      d)   a nucleic acid molecule which anneals under stringent hybridisation conditions to the nucleic
           acid molecule of a), b), or c)
      e)   a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the
           nucleic acid molecule defined in a), b), c) or d).

2.   The isolated nucleic acid molecule according to claim 1, wherein the sequence identity is at least
      80%, preferably at least 95%, especially 100%.

3.   An isolated nucleic acid molecule encoding a hyperimmune serum reactive antigen or a fragment
      thereof comprising a nucleic acid sequence selected from the group consisting of
      a)   a nucleic acid molecule having at least 96% sequence identity to a nucleic acid molecule selected
           from Seq ID No 2-6, 8, 10-16, 18-23, 25-31, 34, 36, 38-42, 44, 47-48, 51, 53, 55-62, 64, 67, 71-76, 78-
           79, 81-94, 96-100.
      b)   a nucleic acid molecule which is complementary to the nucleic acid molecule of a),
      c)   a nucleic acid molecule comprising at least 15 sequential bases of the nucleic acid molecule of a)
           or b)
      d)   a nucleic acid molecule which anneals under stringent hybridisation conditions to the nucleic
           acid molecule of a), b) or c),
      e)   a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the
           nucleic acid defined in a), b), c) or d).

4.   An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group
      consisting of
      a)   a nucleic acid molecule selected from Seq ID No 9, 17, 24, 32, 37, 43, 52, 54, 65-66, 70, 80,
      b)   a nucleic acid molecule which is complementary to the nucleic acid of a),
      c)   a nucleic acid molecule which, but for the degeneracy of the genetic code, would hybridise to the
           nucleic acid defined in a), b), c) or d).

5.   The nucleic acid molecule according to any one of the claims 1, 2, 3 or 4, wherein the nucleic acid is
      DNA.

6.   The nucleic acid molecule according to any one of the claims 1, 2, 3, 4, or 5 wherein the nucleic acid
      is RNA.

7.   An isolated nucleic acid molecule according to any one of claims 1 to 5, wherein the nucleic acid
      molecule is isolated from a genomic DNA, especially from a *S. pneumoniae* genomic DNA.

8.   A vector comprising a nucleic acid molecule according to any one of claims 1 to 7.

9.   A vector according to claim 8, wherein the vector is adapted for recombinant expression of the
      hyperimmune serum reactive antigens or fragment thereof encoded by the nucleic acid molecule
      according to any one of claims 1 to 7.

- 88 -

10.   A host cell comprising the vector according to claim 8 or 9.

11.   A hyperimmune serum-reactive antigen comprising an amino acid sequence being encoded by a
      nucleic acid molecule according to any one of the claims 1, 2, 5, 6 or 7 and fragments thereof,
      wherein the amino acid sequence is selected from the group consisting of Seq ID No 145, 245-288.

12.   A hyperimmune serum-reactive antigen comprising an amino acid sequence being encoded by a
      nucleic acid molecule according to any one of the claims 3, 5, 6, or 7 and fragments thereof,
      wherein the amino acid sueqnece is selected from the group consisting of Seq ID No 146-150, 152,
      154-160, 162-167, 169-175, 178, 180, 182-186, 188, 191-192, 195, 197, 199-206, 208, 211, 215-220, 222-
      223, 225-238, 240-244.

13.   A hyperimmune serum-reactive antigen comprising an amino acid sequence being encoded by a
      nucleic acid molecule according to any one of the claims 4, 5, 6, or 7 and fragments thereof,
      wherein the amino acid sequence is selected from the group consisting of Seq ID No 153, 161, 168,
      176, 181, 187, 196, 198, 209-210, 214, 224.

14.   Fragments of hyperimmune serum-reactive antigens selected from the group consisting of peptides
      comprising amino acid sequences of column "predicted immunogenic aa" and "location of
      identified immunogenic region" of Table 2; the serum reactive epitopes of Table 2, especially
      peptides comprising amino acid 4-11, 35-64, 66-76, 101-108, 111-119 and 57-114 of Seq ID No 145;
      5-27, 32-64, 92-102, 107-113, 119-125, 133-139, 148-162, 177-187, 195-201, 207-214, 241-251, 254-269,
      285-300, 302-309, 317-324, 332-357, 365-404, 411-425, 443-463, 470-477, 479-487, 506-512, 515-520,
      532-547, 556-596, 603-610, 616-622, 624-629, 636-642, 646-665, 667-674, 687-692, 708-720, 734-739,
      752-757, 798-820, 824-851, 856-865 and 732-763 of Seq ID No 146; 14-21, 36-44, 49-66, 102-127, 162-
      167, 177-196, 45-109 and 145-172 of Seq ID No 147; 17-35, 64-75, 81-92, 100-119, 125-172, 174-183,
      214-222, 230-236, 273-282, 287-303, 310-315, 331-340, 392-398, 412-420, 480-505, 515-523, 525-546,
      553-575, 592-598, 603-609, 617-625, 631-639, 644-651, 658-670, 681-687, 691-704, 709-716, 731-736,
      739-744, 750-763, 774-780, 784-791, 799-805, 809-822, 859-870, 880-885, 907-916, 924-941, 943-949,
      973-986, 1010-1016, 1026-1036, 1045-1054, 1057-1062, 1082-1088, 1095-1102, 1109-1120, 1127-1134,
      1140-1146, 1152-1159, 1169-1179, 1187-1196, 1243-1251, 1262-1273, 1279-1292, 1306-1312, 1332-1343,
      1348-1364, 1379-1390, 1412-1420, 1427-1436, 1458-1468, 1483-1503, 1524-1549, 1574-1588, 1614-1619,
      1672-1685, 1697-1707, 1711-1720, 1738-1753, 1781-1787, 1796-1801, 1826-1843, 132-478, 508-592 and
      1753-1810 of Seq ID No 148; 15-43, 49-55, 71-77, 104-110, 123-130, 162-171, 180-192, 199-205, 219-
      227, 246-254, 264-270, 279-287, 293-308, 312-322, 330-342, 349-356, 369-377, 384-394, 401-406, 416-422,
      432-439, 450-460, 464-474, 482-494, 501-508, 521-529, 536-546, 553-558, 568-574, 584-591, 602-612,
      616-626, 634-646, 653-660, 673-681, 688-698, 705-710, 720-726, 736-749, 833-848, 1-199, 200-337, 418-
      494 and 549-647 of Seq ID No 149; 9-30, 65-96, 99-123, 170-178 and 1-128 of Seq ID No 150; 7-32,
      34-41, 96-106, 127-136, 154-163, 188-199, 207-238, 272-279, 306-312, 318-325, 341-347, 353-360, 387-
      393, 399-406, 434-440, 452-503, 575-580, 589-601, 615-620, 635-640, 654-660, 674-680, 696-701, 710-731,
      1-548 and 660-691 of Seq ID No 151; 4-19, 35-44, 48-59, 77-87, 93-99, 106-111, 130-138, 146-161 and
      78-84 of Seq ID No 152; 24-30, 36-43, 64-86, 93-99, 106-130, 132-145, 148-165, 171-177, 189-220, 230-
      249, 251-263, 293-300, 302-312, 323-329, 338-356, 369-379, 390-412 and 179-193 of Seq ID No 153; 30-
      39, 61-67, 74-81, 90-120, 123-145, 154-167, 169-179, 182-197, 200-206, 238-244, 267-272 and 230-265 of
      Seq ID No 154; 14-20, 49-65, 77-86 and 2-68 of Seq ID No 155; 4-9, 26-35, 42-48, 53-61, 63-85, 90-101,
      105-111, 113-121, 129-137, 140-150, 179-188, 199-226, 228-237, 248-255, 259-285, 299-308, 314-331,
      337-343, 353-364, 410-421, 436-442 and 110-144 of Seq ID No 156; 36-47, 55-63, 94-108, 129-134, 144-
      158, 173-187, 196-206, 209-238, 251-266, 270-285, 290-295, 300-306, 333-344, 346-354, 366-397, 404-410,
      422-435, 439-453, 466-473, 515-523, 529-543, 554-569, 571-585, 590-596, 607-618, 627-643, 690-696,
      704-714, 720-728, 741-749, 752-767, 780-799, 225-247 and 480-507 of Seq ID No 157; 16-25, 36-70, 80-

93, 100-106 and 78-130 of Seq ID No 158; 18-27, 41-46, 50-57, 65-71, 79-85, 93-98, 113-128, 144-155, 166-178, 181-188, 201-207, 242-262, 265-273, 281-295, 303-309, 318-327 and 36-64 of Seq ID No 159; 7-29, 31-44, 50-59, 91-96, 146-153, 194-201, 207-212, 232-238, 264-278, 284-290, 296-302, 326-353, 360-370, 378-384, 400-405, 409-418, 420-435, 442-460, 499-506, 529-534, 556-562, 564-576, 644-651, 677-684, 687-698, 736-743, 759-766, 778-784, 808-814, 852-858, 874-896, 920-925, 929-935, 957-965, 1003-1012, 1021-1027, 1030-1044, 1081-1087, 1101-1111, 1116-1124, 1148-1159, 1188-1196, 1235-1251, 1288-1303, 1313-1319, 1328-1335, 1367-1373, 1431-1437, 1451-1458, 1479-1503, 1514-1521, 1530-1540, 1545-1552, 1561-1568, 1598-1605, 1617-1647, 1658-1665, 1670-1676, 1679-1689, 1698-1704, 1707-1713, 1732-1738, 1744-1764, 1-70, 154-189, 922-941, 1445-1462 and 1483-1496 of Seq ID No 160; 6-51, 81-91, 104-113, 126-137, 150-159, 164-174, 197-209, 215-224, 229-235, 256-269, 276-282, 307-313, 317-348, 351-357, 376-397, 418-437, 454-464, 485-490, 498-509, 547-555, 574-586, 602-619 and 452-530 of Seq ID No 161; 25-31, 39-47, 49-56, 99-114, 121-127, 159-186, 228-240, 253-269, 271-279, 303-315, 365-382, 395-405, 414-425, 438-453 and 289-384 of Seq ID No 162; 9-24, 41-47, 49-54, 68-78, 108-114, 117-122, 132-140, 164-169, 179-186, 193-199, 206-213, 244-251, 267-274, 289-294, 309-314, 327-333, 209-249 and 286-336 of Seq ID No 163; 9-28, 53-67, 69-82, 87-93, 109-117, 172-177, 201-207, 220-227, 242-247, 262-268, 305-318, 320-325 and 286-306 of Seq ID No 164; 4-10, 26-39, 47-58, 63-73, 86-96, 98-108, 115-123, 137-143, 148-155, 160-176, 184-189, 194-204, 235-240, 254-259, 272-278 and 199-283 of Seq ID No 165; 4-26, 33-39, 47-53, 59-65, 76-83, 91-97, 104-112, 118-137, 155-160, 167-174, 198-207, 242-268, 273-279, 292-315, 320-332, 345-354, 358-367, 377-394, 403-410, 424-439, 445-451, 453-497, 511-518, 535-570, 573-589, 592-601, 604-610 and 202-242 of Seq ID No 166; 8-30, 36-45, 64-71, 76-82, 97-103, 105-112, 134-151, 161-183, 211-234, 253-268, 270-276, 278-284, 297-305, 309-315, 357-362, 366-372, 375-384, 401-407, 409-416, 441-455, 463-470, 475-480, 490-497, 501-513, 524-537, 552-559, 565-576, 581-590, 592-600, 619-625, 636-644, 646-656 and 316-419 of Seq ID No 167; 4-17, 52-58, 84-99, 102-110, 114-120, 124-135, 143-158, 160-173, 177-196, 201-216, 223-250, 259-267, 269-275 and 1-67 of Seq ID No 168; 6-46, 57-67, 69-80, 82-133, 137-143, 147-168, 182-187, 203-209, 214-229, 233-242, 246-280 and 53-93 of Seq ID No 169; 7-40, 50-56, 81-89, 117-123, 202-209, 213-218, 223-229, 248-261, 264-276, 281-288, 303-308, 313-324, 326-332, 340-346, 353-372, 434-443, 465-474, 514-523, 556-564, 605-616, 620-626, 631-636, 667-683, 685-699, 710-719, 726-732, 751-756, 760-771, 779-788, 815-828, 855-867, 869-879, 897-902, 917-924, 926-931, 936-942, 981-1000, 1006-1015, 1017-1028, 1030-1039, 1046-1054, 1060-1066, 1083-1092, 1099-1112, 1122-1130, 1132-1140, 1148-1158, 1161-1171, 1174-1181, 1209-1230, 1236-1244, 1248-1254, 1256-1267, 1269-1276, 1294-1299, 1316-1328, 1332-1354, 1359-1372, 1374-1380, 1384-1390, 1395-1408, 1419-1425, 1434-1446, 1453-1460, 1465-1471, 1474-1493, 1505-1515, 1523-1537, 1547-1555, 1560-1567, 1577-1605, 1633-1651, 1226-1309, 1455-1536 and 1538-1605 of Seq ID No 170; 4-10, 31-39, 81-88, 106-112, 122-135, 152-158, 177-184, 191-197, 221-227, 230-246, 249-255, 303-311, 317-326, 337-344, 346-362, 365-371, 430-437, 439-446, 453-462, 474-484 and 449-467 of Seq ID No 171; 9-15, 24-35, 47-55, 122-128, 160-177, 188-196, 202-208, 216-228, 250-261, 272-303, 318-324, 327-339, 346-352, 355-361, 368-373, 108-218 and 344-376 of Seq ID No 172; 6-14, 17-48, 55-63, 71-90, 99-109, 116-124, 181-189, 212-223, 232-268, 270-294, 297-304, 319-325, 340-348, 351-370, 372-378, 388-394, 406-415, 421-434 and 177-277 of Seq ID No 173; 21-39, 42-61, 65-75, 79-85, 108-115 and 11-38 of Seq ID No 174; 4-17, 26-39, 61-76, 103-113, 115-122, 136-142, 158-192, 197-203, 208-214, 225-230, 237-251 and 207-225 of Seq ID No 175; 5-11, 27-36, 42-53, 62-70, 74-93, 95-104, 114-119, 127-150, 153-159, 173-179, 184-193, 199-206, 222-241, 248-253, 257-280, 289-295, 313-319, 322-342, 349-365, 368-389, 393-406, 408-413, 426-438, 447-461, 463-470, 476-495, 532-537, 543-550 and 225-246 of Seq ID No 176; 4-29, 68-82, 123-130, 141-147, 149-157, 178-191, 203-215, 269-277, 300-307, 327-335, 359-370, 374-380, 382-388, 393-400, 410-417, 434-442, 483-492, 497-503, 505-513, 533-540, 564-569, 601-607, 639-647, 655-666, 693-706, 712-718, 726-736, 752-758, 763-771, 774-780, 786-799, 806-812, 820-828, 852-863, 884-892, 901-909, 925-932, 943-948, 990-996, 1030-1036, 1051-1059, 1062-1068, 1079-1086, 1105-1113, 1152-1162, 1168-1179, 1183-1191, 1204-1210, 1234-1244, 1286-1295, 1318-1326, 1396-1401, 1451-1460, 1465-1474, 1477-1483, 1488-1494, 1505-1510, 1514-1521, 1552-1565, 1593-1614, 1664-1672, 1677-1685, 1701-1711, 1734-1745, 1758-1770, 1784-1798, 1840-1847, 1852-1873, 1885-1891, 1906-1911, 1931-1939, 1957-1970, 1977-1992, 2014-2020, 2026-2032, 2116-2134, 1-348, 373-490, 573-767, 903-1043, 1155-1198, 1243-1482, 1550-1595, 1682-1719, 1793-1921 and 2008-2110 of Seq ID No 177; 10-35, 39-52, 107-112, 181-188, 226-236, 238-253,

258-268, 275-284, 296-310, 326-338, 345-368, 380-389, 391-408, 410-418, 420-429, 444-456, 489-505, 573-588, 616-623, 637-643, 726-739, 741-767, 785-791, 793-803, 830-847, 867-881, 886-922, 949-956, 961-980, 988-1004, 1009-1018, 1027-1042, 1051-1069, 1076-1089, 1108-1115, 1123-1135, 1140-1151, 1164-1179, 1182-1191, 1210-1221, 1223-1234, 1242-1250, 1255-1267, 1281-1292, 1301-1307, 1315-1340, 1348-1355, 1366-1373, 1381-1413, 1417-1428, 1437-1444, 1453-1463, 1478-1484, 1490-1496, 1498-1503, 1520-1536, 1538-1546, 1548-1570, 1593-1603, 1612-1625, 1635-1649, 1654-1660, 1670-1687, 1693-1700, 1705-1711, 1718-1726, 1729-1763, 1790-1813, 1871-1881, 1893-1900, 1907-1935, 1962-1970, 1992-2000, 2006-2013, 2033-2039, 2045-2051, 2055-2067, 2070-2095, 2097-2110, 2115-2121, 2150-2171, 2174-2180, 2197-2202, 2206-2228 and 1526-1560 of Seq ID No 178; 4-17, 35-48, 54-76, 78-107, 109-115, 118-127, 134-140, 145-156, 169-174, 217-226, 232-240, 256-262, 267-273, 316-328, 340-346, 353-360, 402-409, 416-439, 448-456, 506-531, 540-546, 570-578, 586-593, 595-600, 623-632, 662-667, 674-681, 689-705, 713-724, 730-740, 757-763, 773-778, 783-796, 829-835, 861-871, 888-899, 907-939, 941-955, 957-969, 986-1000, 1022-1028, 1036-1044, 1068-1084, 1095-1102, 1118-1124, 1140-1146, 1148-1154, 1168-1181, 1185-1190, 1197-1207, 1218-1226, 1250-1270, 1272-1281, 1284-1296, 1312-1319, 1351-1358, 1383-1409, 1422-1428, 1438-1447, 1449-1461, 1482-1489, 1504-1510, 1518-1527, 1529-1537, 1544-1551, 1569-1575, 1622-1628, 1631-1637, 1682-1689, 1711-1718, 1733-1740, 1772-1783, 1818-1834, 1859-1872, 1-64 and 128-495 of Seq ID No 179; 8-28, 32-37, 62-69, 119-125, 137-149, 159-164, 173-189, 200-205, 221-229, 240-245, 258-265, 268-276, 287-293, 296-302, 323-329 and 1-95 of Seq ID No 180; 9-18, 25-38, 49-63, 65-72, 74-81, 94-117, 131-137, 139-146, 149-158, 162-188, 191-207, 217-225, 237-252, 255-269, 281-293, 301-326, 332-342, 347-354, 363-370, 373-380, 391-400, 415-424, 441-447 and 75-107 of Seq ID No 181; 4-24, 64-71, 81-87, 96-116, 121-128, 130-139, 148-155, 166-173, 176-184, 203-215, 231-238, 243-248, 256-261, 280-286, 288-306, 314-329 and 67-148 of Seq ID No 182; 4-10, 19-37, 46-52, 62-81, 83-89, 115-120, 134-139, 141-151, 168-186, 197-205, 209-234, 241-252, 322-335, 339-345, 363-379, 385-393, 403-431, 434-442, 447-454, 459-465, 479-484, 487-496 and 404-420 of Seq ID No 183; 10-35, 46-66, 71-77, 84-93, 96-122, 138-148, 154-172, 182-213, 221-233, 245-263, 269-275, 295-301, 303-309, 311-320, 324-336, 340-348, 351-359, 375-381 and 111-198 of Seq ID No 184; 14-25, 30-42, 47-61, 67-75, 81-91, 98-106, 114-122, 124-135, 148-193, 209-227 and 198-213 of Seq ID No 185; 5-18, 45-50, 82-90, 97-114, 116-136, 153-161, 163-171, 212-219, 221-227, 240-249, 267-281, 311-317, 328-337, 375-381, 390-395, 430-436, 449-455, 484-495, 538-543, 548-554, 556-564, 580-586, 596-602 and 493-606 of Seq ID No 186; 9-25, 28-34, 37-44, 61-68, 75-81, 88-96, 98-111, 119-133, 138-150, 152-163, 168-182, 186-194, 200-205, 216-223, 236-245, 257-264, 279-287, 293-304, 311-318, 325-330, 340-346, 353-358, 365-379, 399-409, 444-453 and 303-391 of Seq ID No 187; 16-36, 55-61, 66-76, 78-102, 121-130, 134-146, 150-212, 221-239, 255-276, 289-322, 329-357 and 29-59 of Seq ID No 188; 8-27, 68-74, 77-99, 110-116, 124-141, 171-177, 202-217, 221-228, 259-265, 275-290, 293-303, 309-325, 335-343, 345-351, 365-379, 384-394, 406-414, 423-437, 452-465, 478-507, 525-534, 554-560, 611-624, 628-651, 669-682, 742-747, 767-778, 782-792, 804-812, 820-836, 79-231 and 359-451 of Seq ID No 189; 5-28, 39-45, 56-62, 67-74, 77-99, 110-117, 124-141, 168-176, 200-230, 237-244, 268-279, 287-299, 304-326, 329-335, 348-362, 370-376, 379-384, 390-406, 420-429, 466-471, 479-489, 495-504, 529-541, 545-553, 561-577, 598-604, 622-630, 637-658, 672-680, 682-688, 690-696, 698-709, 712-719, 724-736, 738-746, 759-769, 780-786, 796-804, 813-818, 860-877, 895-904, 981-997, 1000-1014, 1021-1029, 1-162, 206-224, 254-350, 414-514 and 864-938 of Seq ID No 190; 4-11, 19-49, 56-66, 68-101, 109-116, 123-145, 156-165, 177-185, 204-221, 226-234, 242-248, 251-256, 259-265, 282-302, 307-330, 340-349, 355-374, 377-383, 392-400, 422-428, 434-442, 462-474 and 266-322 of Seq ID No 191; 14-43, 45-57, 64-74, 80-87, 106-127, 131-142, 145-161, 173-180, 182-188, 203-210, 213-219, 221-243, 245-254, 304-311, 314-320, 342-348, 354-365, 372-378, 394-399, 407-431, 436-448, 459-465, 470-477, 484-490, 504-509, 531-537, 590-596, 611-617, 642-647, 723-734, 740-751, 754-762, 764-774, 782-797, 807-812, 824-831, 838-845, 877-885, 892-898, 900-906, 924-935, 940-946, 982-996, 1006-1016, 1033-1043, 1051-1056, 1058-1066, 1094-1108, 1119-1126, 1129-1140, 1150-1157, 1167-1174, 1176-1185, 1188-1201, 1209-1216, 1220-1228, 1231-1237, 1243-1248, 1253-1285, 1288-1297, 1299-1307, 1316-1334, 1336-1343, 1350-1359, 1365-1381, 1390-1396, 1412-1420, 1427-1439, 1452-1459, 1477-1484, 1493-1512, 1554-1559, 1570-1578, 1603-1608, 1623-1630, 1654-1659, 1672-1680, 1689-1696, 1705-1711, 1721-1738, 1752-1757, 1773-1780, 1817-1829, 1844-1851, 1856-1863, 1883-1895, 1950-1958, 1974-1990, 172-354, 384-448, 464-644, 648-728 and 1357-1370 of Seq ID No 192; 8-27, 68-74, 77-99, 110-116, 124-141, 169-

176, 201-216, 220-227, 258-264, 274-289, 292-302, 308-324, 334-342, 344-350, 364-372, 377-387, 399-407, 416-429, 445-458, 471-481, 483-500, 518-527, 547-553, 604-617, 621-644, 662-675, 767-778, 809-816, 15-307, 350-448 and 496-620 of Seq ID No 193; 4-17, 24-29, 53-59, 62-84, 109-126, 159-164, 189-204, 208-219, 244-249, 274-290, 292-302, 308-324, 334-342, 344-350, 378-389, 391-397, 401-409, 424-432, 447-460, 470-479, 490-504, 521-529, 538-544, 549-555, 570-577, 583-592, 602-608, 615-630, 635-647, 664-677, 692-698, 722-731, 733-751, 782-790, 793-799, 56-267, 337-426 and 495-601 of Seq ID No 194; 12-22, 49-59, 77-89, 111-121, 136-148, 177-186, 207-213, 217-225, 227-253, 259-274, 296-302, 328-333, 343-354, 374-383, 424-446, 448-457, 468-480, 488-502, 507-522, 544-550, 553-560, 564-572, 587-596, 604-614, 619-625, 629-635, 638-656, 662-676, 680-692, 697-713, 720-738, 779-786, 833-847, 861-869, 880-895, 897-902, 911-917, 946-951, 959-967, 984-990, 992-1004, 1021-1040, 1057-1067, 1073-1080 and 381-403 of Seq ID No 195; 4-10, 26-31, 46-56, 60-66, 70-79, 86-94, 96-102, 109-118, 132-152, 164-187, 193-206, 217-224 and 81-149 of Seq ID No 196; 4-21, 26-37, 48-60, 71-82, 109-117, 120-128, 130-136, 142-147, 181-187, 203-211, 216-223, 247-255, 257-284, 316-325, 373-379, 395-400, 423-435, 448-456, 479-489, 512-576, 596-625, 641-678, 680-688, 692-715 and 346-453 of Seq ID No 197; 10-16, 25-31, 34-56, 58-69, 71-89, 94-110, 133-176, 186-193, 208-225, 240-250, 259-266, 302-307, 335-341, 376-383, 410-416 and 316-407 of Seq ID No 198; 11-29, 42-56, 60-75, 82-88, 95-110, 116-126, 132-143, 145-160, 166-172, 184-216 and 123-164 of Seq ID No 199; 11-29, 54-63, 110-117, 139-152, 158-166, 172-180, 186-193, 215-236, 240-251, 302-323, 330-335, 340-347, 350-366, 374-381 and 252-299 of Seq ID No 200; 18-27, 35-42, 50-56, 67-74, 112-136, 141-153, 163-171, 176-189, 205-213, 225-234, 241-247, 253-258, 269-281, 288-298, 306-324, 326-334, 355-369, 380-387 and 289-320 of Seq ID No 201; 7-15, 19-41, 56-72, 91-112, 114-122, 139-147, 163-183, 196-209, 258-280, 326-338, 357-363, 391-403, 406-416 and 360-378 of Seq ID No 202; 11-18, 29-41, 43-49, 95-108, 142-194, 204-212, 216-242, 247-256, 264-273 and 136-149 of Seq ID No 203; 18-24, 33-40, 65-79, 89-102, 113-119, 130-137, 155-161, 173-179, 183-203, 205-219, 223-231, 245-261, 267-274, 296-306, 311-321, 330-341, 344-363, 369-381, 401-408, 415-427, 437-444, 453-464, 472-478, 484-508, 517-524, 526-532, 543-548 and 59-180 of Seq ID No 204; 5-13, 52-65, 67-73, 97-110, 112-119, 134-155 and 45-177 of Seq ID No 205; 6-28, 34-43, 57-67, 75-81, 111-128, 132-147, 155-163, 165-176, 184-194, 208-216, 218-229, 239-252, 271-278, 328-334, 363-376, 381-388, 426-473, 481-488, 492-498, 507-513, 536-546, 564-582, 590-601, 607-623, 148-269, 420-450 and 610-648 of Seq ID No 206; 4-12, 20-38, 69-75, 83-88, 123-128, 145-152, 154-161, 183-188, 200-213, 245-250, 266-272, 306-312, 332-339, 357-369, 383-389, 395-402, 437-453, 455-470, 497-503 and 1-112 of Seq ID No 207; 35-59, 74-86, 111-117, 122-137 and 70-154 of Seq ID No 208; 26-42, 54-61, 65-75, 101-107, 123-130, 137-144, 148-156, 164-172, 177-192, 213-221, 231-258 and 157-249 of Seq ID No 209; 29-38, 61-67, 77-87, 94-100, 105-111, 118-158 and 1-97 of Seq ID No 210; 7-21, 30-48, 51-58, 60-85, 94-123, 134-156, 160-167, 169-183, 186-191, 216-229, 237-251, 257-267, 272-282, 287-298 and 220-243 of Seq ID No 211; 6-29, 34-47, 56-65, 69-76, 83-90, 123-134, 143-151, 158-178, 197-203, 217-235, 243-263, 303-309, 320-333, 338-348, 367-373, 387-393, 407-414, 416-427, 441-457, 473-482, 487-499, 501-509, 514-520, 530-535, 577-583, 590-602, 605-612, 622-629, 641-670, 678-690, 37-71 and 238-307 of Seq ID No 212; 7-40, 121-132, 148-161, 196-202, 209-215, 221-235, 248-255, 271-280, 288-295, 330-339, 395-409, 414-420, 446-451, 475-487, 556-563, 568-575, 580-586, 588-595, 633-638, 643-648, 652-659, 672-685, 695-700, 710-716, 737-742, 749-754, 761-767, 775-781, 796-806, 823-835, 850-863, 884-890, 892-900, 902-915, 934-941 and 406-521 of Seq ID No 213; 9-18, 24-46, 51-58, 67-77, 85-108, 114-126, 129-137, 139-146, 152-165, 173-182, 188-195, 197-204, 217-250, 260-274, 296-313, 343-366, 368-384, 427-434, 437-446, 449-455, 478-484, 492-506, 522-527, 562-591, 599-606, 609-618, 625-631, 645-652 and 577-654 of Seq ID No 214; 13-20, 26-37, 41-53, 56-65, 81-100, 102-114, 118-127, 163-188, 196-202, 231-238, 245-252, 266-285, 293-298, 301-306 and 19-78 of Seq ID No 215; 10-23, 32-42, 54-66, 73-91, 106-113, 118-127, 139-152, 164-173, 198-207, 210-245, 284-300, 313-318, 330-337, 339-346, 354-361, 387-393, 404-426, 429-439, 441-453, 467-473, 479-485, 496-509, 536-544, 551-558, 560-566, 569-574, 578-588, 610-615, 627-635, 649-675, 679-690, 698-716, 722-734, 743-754, 769-780, 782-787 and 480-550 of Seq ID No 216; 6-39, 42-50, 60-68, 76-83, 114-129, 147-162, 170-189, 197-205, 217-231, 239-248, 299-305, 338-344, 352-357, 371-377, 380-451, 459-483, 491-499, 507-523, 537-559, 587-613, 625-681, 689-729, 737-781, 785-809, 817-865, 873-881, 889-939, 951-975, 983-1027, 1031-1055, 1063-1071, 1079-1099, 1103-1127, 1151-1185, 1197-1261, 1269-1309, 1317-1333, 1341-1349, 1357-1465, 1469-1513, 1517-1553, 1557-1629, 1637-1669, 1677-1701, 1709-1725,

1733-1795, 1823-1849, 1861-1925, 1933-1973, 1981-2025, 2029-2053, 2061-2109, 2117-2125, 2133-2183, 2195-2219, 2227-2271, 2275-2299, 2307-2315, 2323-2343, 2347-2371, 2395-2429, 2441-2529, 2537-2569, 2577-2601, 2609-2625, 2633-2695, 2699-2737, 2765-2791, 2803-2867, 2889-2913, 2921-2937, 2945-2969, 2977-2985, 2993-3009, 3023-3045, 3073-3099, 3111-3167, 3175-3215, 3223-3267, 3271-3295, 3303-3351, 3359-3367, 3375-3425, 3437-3461, 3469-3513, 3517-3541, 3549-3557, 3565-3585, 3589-3613, 3637-3671, 3683-3747, 3755-3795, 3803-3819, 3827-3835, 3843-3951, 3955-3999, 4003-4039, 4043-4115, 4123-4143, 4147-4171, 4195-4229, 4241-4305, 4313-4353, 4361-4377, 4385-4393, 4401-4509, 4513-4557, 4561-4597, 4601-4718, 4749-4768, 74-171, 452-559 and 2951-3061 of Seq ID No 217; 16-22, 30-51, 70-111, 117-130, 137-150, 171-178, 180-188, 191-196 and 148-181 of Seq ID No 218; 6-19, 21-46, 50-56, 80-86, 118-126, 167-186, 189-205, 211-242, 244-267, 273-286, 290-297, 307-316, 320-341 and 34-60 of Seq ID No 219; 5-26, 33-43, 48-54, 58-63, 78-83, 113-120, 122-128, 143-152, 157-175, 185-192, 211-225, 227-234, 244-256, 270-281, 284-290, 304-310, 330-337, 348-355, 362-379, 384-394, 429-445, 450-474, 483-490, 511-520, 537-546, 548-554, 561-586, 590-604, 613-629, 149-186, 285-431 and 573-659 of Seq ID No 220; 5-26, 49-59, 61-67, 83-91, 102-111, 145-157, 185-192, 267-272, 279-286, 292-298, 306-312, 134-220, 235-251 and 254-280 of Seq ID No 221; 5-19, 72-79, 83-92, 119-124, 140-145, 160-165, 167-182, 224-232, 240-252, 259-270, 301-310, 313-322, 332-343, 347-367, 384-398, 416-429, 431-446, 454-461 and 1-169 of Seq ID No 222; 8-17, 26-31, 56-62, 75-83, 93-103, 125-131, 135-141, 150-194, 205-217, 233-258, 262-268, 281-286 and 127-168 of Seq ID No 223; 6-12, 69-75, 108-115, 139-159, 176-182, 194-214 and 46-161 of Seq ID No 224; 6-13, 18-27, 39-48, 51-59, 66-73, 79-85, 95-101, 109-116, 118-124, 144-164, 166-177, 183-193, 197-204, 215-223, 227-236, 242-249, 252-259, 261-270, 289-301, 318-325 and 12-58 of Seq ID No 225; 4-10, 26-32, 48-60, 97-105, 117-132, 138-163, 169-185, 192-214, 219-231, 249-261, 264-270, 292-308, 343-356, 385-392, 398-404, 408-417, 435-441 and 24-50 of Seq ID No 226; 10-40, 42-48, 51-61, 119-126 and 1-118 of Seq ID No 227; 5-17, 40-58, 71-83, 103-111, 123-140, 167-177, 188-204 and 116-128 of Seq ID No 228; 4-9, 11-50, 57-70, 112-123, 127-138 and 64-107 of Seq ID No 229; 9-39, 51-67 and 1-101 of Seq ID No 230; 5-14, 17-25, 28-46, 52-59, 85-93, 99-104, 111-120, 122-131, 140-148, 158-179, 187-197, 204-225, 271-283, 285-293 and 139-155 of Seq ID No 231; 42-70, 73-90, 92-108, 112-127, 152-164, 166-172, 181-199, 201-210, 219-228, 247-274, 295-302, 322-334, 336-346, 353-358, 396-414, 419-425, 432-438, 462-471, 518-523, 531-536, 561-567, 576-589, 594-612, 620-631, 665-671, 697-710, 718-731, 736-756, 765-771, 784-801 and 626-653 of Seq ID No 232; 8-28, 41-51, 53-62, 68-74, 79-85, 94-100, 102-108, 114-120, 130-154, 156-162, 175-180, 198-204, 206-213, 281-294, 308-318, 321-339, 362-368, 381-386, 393-399, 407-415 and 2-13 of Seq ID No 233; 4-39, 48-65, 93-98, 106-112, 116-129 and 10-36 of Seq ID No 234; 25-32, 35-50, 66-71, 75-86, 90-96, 123-136, 141-151, 160-179, 190-196, 209-215, 222-228, 235-242, 257-263, 270-280 and 209-247 of Seq ID No 235; 5-29, 31-38, 50-57, 62-75, 83-110, 115-132, 168-195, 197-206, 216-242, 249-258, 262-269, 333-340, 342-350, 363-368, 376-392, 400-406, 410-421, 423-430, 436-442, 448-454, 460-466, 471-476, 491-496, 511-516, 531-536, 551-556, 571-576, 585-591, 599-605, 27-70, 219-293, 441-504 and 512-584 of Seq ID No 236; 4-12, 14-34, 47-75, 83-104, 107-115, 133-140, 148-185, 187-196, 207-212, 224-256, 258-265, 281-287, 289-296, 298-308, 325-333, 345-355, 365-371, 382-395, 424-435, 441-457, 465-472, 483-491, 493-505, 528-534, 536-546, 552-558, 575-584, 589-600, 616-623 and 576-591 of Seq ID No 237; 4-76, 78-89, 91-126, 142-148, 151-191, 195-208, 211-223, 226-240, 256-277, 279-285, 290-314, 317-323, 358-377, 381-387, 391-396, 398-411, 415-434, 436-446, 454-484, 494-512, 516-523, 538-552, 559-566, 571-577, 579-596, 599-615, 620-627, 635-644, 694-707, 720-734, 737-759, 761-771 and 313-329 of Seq ID No 238; 7-38, 44-49, 79-89, 99-108, 117-123, 125-132, 137-146, 178-187, 207-237, 245-255, 322-337, 365-387, 398-408, 445-462, 603-608, 623-628, 644-650, 657-671, 673-679 and 111-566 of Seq ID No 239; 6-20, 22-35, 39-45, 58-64, 77-117, 137-144, 158-163, 205-210, 218-224, 229-236, 239-251, 263-277, 299-307, 323-334, 353-384, 388-396, 399-438, 443-448, 458-463, 467-478, 481-495, 503-509, 511-526, 559-576, 595-600, 612-645, 711-721, 723-738, 744-758, 778-807 and 686-720 of Seq ID No 240; 10-33, 35-41, 72-84, 129-138, 158-163, 203-226, 243-252, 258-264, 279-302, 322-329, 381-386, 401-406, 414-435 and 184-385 of Seq ID No 241; 4-9, 19-24, 41-47, 75-85, 105-110, 113-146 and 45-62 of Seq ID No 242; 4-25, 52-67, 117-124, 131-146, 173-180, 182-191, 195-206, 215-221, 229-236, 245-252, 258-279, 286-291, 293-302, 314-320, 327-336, 341-353, 355-361, 383-389 and 1-285 of Seq ID No 243; 14-32, 38-50, 73-84, 93-105, 109-114 and 40-70 of Seq ID No 244; 5-26 and 22-34 of Seq ID No 245; 23-28 and 13-39 of Seq ID No 246; 8-14 and 21-34 of

Seq ID No 247; 4-13, 20-29, 44-50, 59-74 and 41-69 of Seq ID No 248; 4-9, 19-42, 48-59, 71-83 and 57-91 of Seq ID No 249; 4-14 and 10-28 of Seq ID No 250; 22-28, 32-42, 63-71, 81-111, 149-156, 158-167, 172-180, 182-203, 219-229 and 27-49 of Seq ID No 251; 17-27 and 23-32 of Seq ID No 252; 18-24 and 28-38 of Seq ID No 253; 9-15 and 13-27 of Seq ID No 254; 13-22 and 18-29 of Seq ID No 255; 17-26 and 2-11 of Seq ID No 256; 4-33 and 16-32 of Seq ID No 257; 4-10, 37-43, 54-84, 92-127 and 15-62 of Seq ID No 258; 4-14, 20-32, 35-60, 69-75, 79-99, 101-109, 116-140 and 124-136 of Seq ID No 259; 2-13 of Seq ID No 260; 4-13, 28-42 and 42-57 of Seq ID No 261; 4-14, 27-44 and 14-35 of Seq ID No 262; 4-12 and 1-27 of Seq ID No 263; 4-18, 39-45, 47-74 and 35-66 of Seq ID No 264; 8-20, 43-77 and 17-36 of Seq ID No 265; 4-30, 35-45, 51-57 and 35-49 of Seq ID No 266; 4-24, 49-57 and 15-34 of Seq ID No 267; 4-22 and 8-27 of Seq ID No 268; 13-25, 32-59, 66-80 and 21-55 of Seq ID No 269; 4-10, 24-33, 35-42, 54-65, 72-82, 98-108 and 15-30 of Seq ID No 270; 8-19 and 17-47 of Seq ID No 271; 12-18, 40-46 and 31-52 of Seq ID No 272; 4-20, 35-78, 83-102, 109-122 and 74-86 of Seq ID No 273; 7-17, 21-41, 46-63 and 2-20 of Seq ID No 274; 30-37 and 2-33 of Seq ID No 275; 4-13, 17-25 and 1-15 of Seq ID No 276; 17-31, 44-51 and 20-51 of Seq ID No 277; 20-30 and 5-23 of Seq ID No 278; 13-33, 48-71 and 92-110 of Seq ID No 279; 4-9, 50-69, 76-88, 96-106, 113-118 and 12-34 of Seq ID No 280; 4-24 and 6-26 of Seq ID No 281; 7-26 and 14-30 of Seq ID No 282; 9-39, 46-68, 75-82, 84-103 and 26-44 of Seq ID No 283; 4-30, 33-107 and 58-84 of Seq ID No 284; 4-12 and 9-51 of Seq ID No 285; 12-18, 29-37 and 6-37 of Seq ID No 286; 4-21, 33-52, 64-71 and 16-37 of Seq ID No 287; 9-19 and 2-30 of Seq ID No 288; 20-37 of Seq ID No 245; 8-27 of Seq ID No 246; 10-27 of Seq ID No 247; 42-59 and 52-69 of Seq ID No 248; 63-80 and 74-91 of Seq ID No 249; 11-28 of Seq ID No 250; 28-49 of Seq ID No 251; 15-32 of Seq ID No 252; 4-20 of Seq ID No 253; 10-27 of Seq ID No 254; 17-34 of Seq ID No 255; 1-18 of Seq ID No 256; 16-33 of Seq ID No 257; 16-36, 30-49 and 43-62 of Seq ID No 258; 122-139 of Seq ID No 259; 1-18 of Seq ID No 260; 41-58 of Seq ID No 261; 15-35 of Seq ID No 262; 2-27 of Seq ID No 263; 18-36 of Seq ID No 265; 34-51 of Seq ID No 266; 9-27 of Seq ID No 268; 22-47 of Seq ID No 269; 18-36 and 29-47 of Seq ID No 271; 32-52 of Seq ID No 272; 72-89 of Seq ID No 273; 3-20 of Seq ID No 274; 3-21 and 15-33 of Seq ID No 275; 1-18 of Seq ID No 276; 6-23 of Seq ID No 278; 93-110 of Seq ID No 279; 13-34 of Seq ID No 280; 7-26 and 9-26 of Seq ID No 281; 16-33 of Seq ID No 282; 27-44 of Seq ID No 283; 67-84 of Seq ID No 284; 10-33 and 26-50 of Seq ID No 285; 7-25 and 19-37 of Seq ID No 286; 17-37 of Seq ID No 287; 3-20 and 13-30 of Seq ID No 288; 62-80 and 75-93 of Seq ID No 145; 92-108 of Seq ID No 147; 332-349, 177-200 and 1755-1777 of Seq ID No 148; 109-133, 149-174, 260-285 and 460-485 of Seq ID No 149; 26-47 and 42-64 of Seq ID No 150; 22-41, 35-54, 115-130, 306-325, 401-420 and 454-478 of Seq ID No 151; 22-45 of Seq ID No 155; 156-174, 924-940, 1485-1496, 1447-1462 and 1483-1498 of Seq ID No 160; 457-475 of Seq ID No 161; 302-325 of Seq ID No 163; 288-305 of Seq ID No 164; 244-266 and 260-282 of Seq ID No 165; 204-225 and 220-241 of Seq ID No 166; 324-345, 340-361, 356-377, 372-393 and 388-408 of Seq ID No 167; 39-64 of Seq ID No 168; 54-76 and 70-92 of Seq ID No 169; 1227-1247, 1539-1559, 1554-1574, 1569-1589, 1584-1604, 1242-1262, 1272-1292, 1287-1308, 1456-1477, 1472-1494, 1488-1510 and 1505-1526 of Seq ID No 170; 351-368 of Seq ID No 172; 179-200, 195-216, 211-232, 227-248 and 243-263 of Seq ID No 173; 13-37 of Seq ID No 174; 208-224 of Seq ID No 175; 42-64, 59-81, 304-328, 323-348, 465-489, 968-992, 1399-1418, 1412-1431 and 2092-2111 of Seq ID No 177; 1528-1547 and 1541-1560 of Seq ID No 178; 184-200, 367-388, 382-403, 409-429, 425-444 and 438-457 of Seq ID No 179; 27-50 and 45-67 of Seq ID No 180; 114-131 and 405-419 of Seq ID No 183; 113-134, 129-150, 145-166, 161-182 and 177-198 of Seq ID No 184; 495-515 of Seq ID No 186; 346-358 of Seq ID No 187; 208-224 of Seq ID No 190; 178-194, 202-223, 217- 238, 288-308 and 1355-1372 of Seq ID No 192; 57-78 of Seq ID No 194; 347-369, 364-386, 381-403, 398-420, 415-437 and 432-452 of Seq ID No 197; 347-372 of Seq ID No 198; 147-163 of Seq ID No 199; 263-288 of Seq ID No 200; 361-377 of Seq ID No 202; 82-104, 99-121, 116-138, 133-155 and 150-171 of Seq ID No 204; 110-130 and 125-145 of Seq ID No 205; 613-631, 626-644 and 196-213 of Seq ID No 206; 78-100, 95-117, 112-134 and 129-151 of Seq ID No 208; 158-180, 175-197, 192-214, 209-231 and 226-248 of Seq ID No 209; 30-50, 45-65 and 60-79 of Seq ID No 210; 431-455 and 450-474 of Seq ID No 213; 579-601, 596-618, 613-635 and 630-653 of Seq ID No 214; 920-927, 98-119, 114-135, 130-151, 146-167 and 162-182 of Seq ID No 217; 36-59 of Seq ID No 219; 194-216 and 381-404 of Seq ID No 220; 236-251 and 255-279 of Seq ID No 221; 80-100 and

141-164 of Seq ID No 222; 128-154 of Seq ID No 223; 82-100, 95-116 and 111-134 of Seq ID No 224; 55-76, 71-92 and 87-110 of Seq ID No 227; 91-106 of Seq ID No 229; 74-96 of Seq ID No 230; 140-157 of Seq ID No 231; 4-13 of Seq ID No 233; 41-65 and 499-523 of Seq ID No 236; 122-146, 191-215, 288-313, 445-469 and 511-535 of Seq ID No 239; 347-368 of Seq ID No 241; 46-61 of Seq ID No 242; 15-37, 32-57, 101-121, 115-135, 138-158, 152-172, 220-242 and 236-258 of Seq ID No 243.

15.   A process for producing a *S. pneumoniae* hyperimmune serum reactive antigen or a fragment thereof according to any one of the claims 11 to 14 comprising expressing the nucleic acid molecule according to any one of claims 1 to 7.

16.   A process for producing a cell, which expresses a *S. pneumoniae* hyperimmune serum reactive antigen or a fragment thereof according to any one of the claims 11 to 14 comprising transforming or transfecting a suitable host cell with the vector according to claim 8 or claim 9.

17.   A pharmaceutical composition, especially a vaccine, comprising a hyperimmune serum-reactive antigen or a fragment thereof, as defined in any one of claims 11 to 14 or a nucleic acid molecule according to any one of claims 1 to 7.

18.   A pharmaceutical composition, especially a vaccine, according to claim 17, characterized in that it further comprises an immunostimulatory substance, preferably selected from the group comprising polycationic polymers, especially polycationic peptides, immunostimulatory deoxynucleotides (ODNs), peptides containing at least two LysLeuLys motifs, neuroactive compounds, especially human growth hormone, alumn, Freund's complete or incomplete adjuvants or combinations thereof.

19.   Use of a nucleic acid molecule according to any one of claims 1 to 7 or a hyperimmune serum-reactive antigen or fragment thereof according to any one of claims 11 to 14 for the manufacture of a pharmaceutical preparation, especially for the manufacture of a vaccine against *S. pneumoniae* infection.

20.   An antibody, or at least an effective part thereof, which binds at least to a selective part of the hyperimmune serum-reactive antigen or a fragment thereof according to any one of claims 11 to 14.

21.   An antibody according to claim 20, wherein the antibody is a monoclonal antibody.

22.   An antibody according to claim 20 or 21, wherein said effective part comprises Fab fragments.

23.   An antibody according to any one of claims 20 to 22, wherein the antibody is a chimeric antibody.

24.   An antibody according to any one of claims 20 to 23, wherein the antibody is a humanized antibody.

25.   A hybridoma cell line, which produces an antibody according to any one of claims 20 to 24.

26.   A method for producing an antibody according to claim 20, characterized by the following steps:
- initiating an immune response in a non-human animal by administrating an hyperimmune serum-reactive antigen or a fragment thereof, as defined in any one of the claims 11 to 14, to said animal,
- removing an antibody containing body fluid from said animal, and
- producing the antibody by subjecting said antibody containing body fluid to further purification steps.

27. Method for producing an antibody according to claim 21, characterized by the following steps:
    - initiating an immune response in a non-human animal by administrating an hyperimmune serum-reactive antigen or a fragment thereof, as defined in any one of the claims 12 to 15, to said animal,
    - removing the spleen or spleen cells from said animal,
    - producing hybridoma cells of said spleen or spleen cells,
    - selecting and cloning hybridoma cells specific for said hyperimmune serum-reactive antigens or a fragment thereof,
    - producing the antibody by cultivation of said cloned hybridoma cells and optionally further purification steps.

28. Use of the antibodies according to any one of claims 20 to 24 for the preparation of a medicament for treating or preventing *S. pneumoniae* infections.

29. An antagonist, which binds to the hyperimmune serum-reactive antigen or a fragment thereof according to any one of claims 11 to 14.

30. A method for identifying an antagonist capable of binding to the hyperimmune serum-reactive antigen or fragment thereof according to any one of claims 11 to 14 comprising:
    a) contacting an isolated or immobilized hyperimmune serum-reactive antigen or a fragment thereof according to any one of claims 11 to 14 with a candidate antagonist under conditions to permit binding of said candidate antagonist to said hyperimmune serum-reactive antigen or fragment, in the presence of a component capable of providing a detectable signal in response to the binding of the candidate antagonist to said hyperimmune serum reactive antigen or fragment thereof; and
    b) detecting the presence or absence of a signal generated in response to the binding of the antagonist to the hyperimmune serum reactive antigen or the fragment thereof.

31. A method for identifying an antagonist capable of reducing or inhibiting the interaction activity of a hyperimmune serum-reactive antigen or a fragment thereof according to any one of claims 11 to 14 to its interaction partner comprising:
    a) providing a hyperimmune serum reactive antigen or a hyperimmune fragment thereof according to any one of claims 11-14,
    b) providing an interaction partner to said hyperimmune serum reactive antigen or a fragment thereof, especially an antibody according to any one of the claims 20 to 24,
    c) allowing interaction of said hyperimmune serum reactive antigen or fragment thereof to said interaction partner to form a interaction complex,
    d) providing a candidate antagonist,
    e) allowing a competition reaction to occur between the candidate antagonist and the interaction complex,
    f) determining whether the candidate antagonist inhibits or reduces the interaction activities of the hyperimmune serum reactive antigen or the fragment thereof with the interaction partner.

32. Use of any of the hyperimmune serum reactive antigen or fragment thereof according to any one of claims 11 to 14 for the isolation and/or purification and/or identification of an interaction partner of said hyperimmune serum reactive antigen or fragment thereof.

33. A process for *in vitro* diagnosing a disease related to expression of the hyperimmune serum-reactive antigen or a fragment thereof according to any one of claims 11 to 14 comprising determining the presence of a nucleic acid sequence encoding said hyperimmune serum reactive antigen and fragment according to any one of claims 1 to 7 or the presence of the hyperimmune

serum reactive antigen or fragment thereof according to any one of claims 11-14.

34.   A process for *in vitro* diagnosis of a bacterial infection, especially a *S. pneumoniae* infection, comprising analysing for the presence of a nucleic acid sequence encoding said hyperimmune serum reactive antigen and fragment according to any one of claims 1 to 7 or the presence of the hyperimmune serum reactive antigen or fragment thereof according to any one of claims 11 to 14.

35.   Use of the hyperimmune serum reactive antigen or fragment thereof according to any one of claims 11 to 14 for the generation of a peptide binding to said hyperimmune serum reactive antigen or fragment thereof, wherein the peptide is selected from the group comprising anticalines.

36.   Use of the hyperimmune serum-reactive antigen or fragment thereof according to any one of claims 11 to 14 for the manufacture of a functional nucleic acid, wherein the functional nucleic acid is selected from the group comprising aptamers and spiegelmers.

37.   Use of a nucleic acid molecule according to any one of claims 11 to 14 for the manufacture of a functional ribonucleic acid, wherein the functional ribonucleic acid is selected from the group comprising ribozymes, antisense nucleic acids and siRNA.

Fig. 1

Fig.1

Fig.1

A

B

Fig. 2

A

Cfu   1st round of MACS with PSPn3-IgG

cfu   2st round of MACS with PSPn3-IgG

B
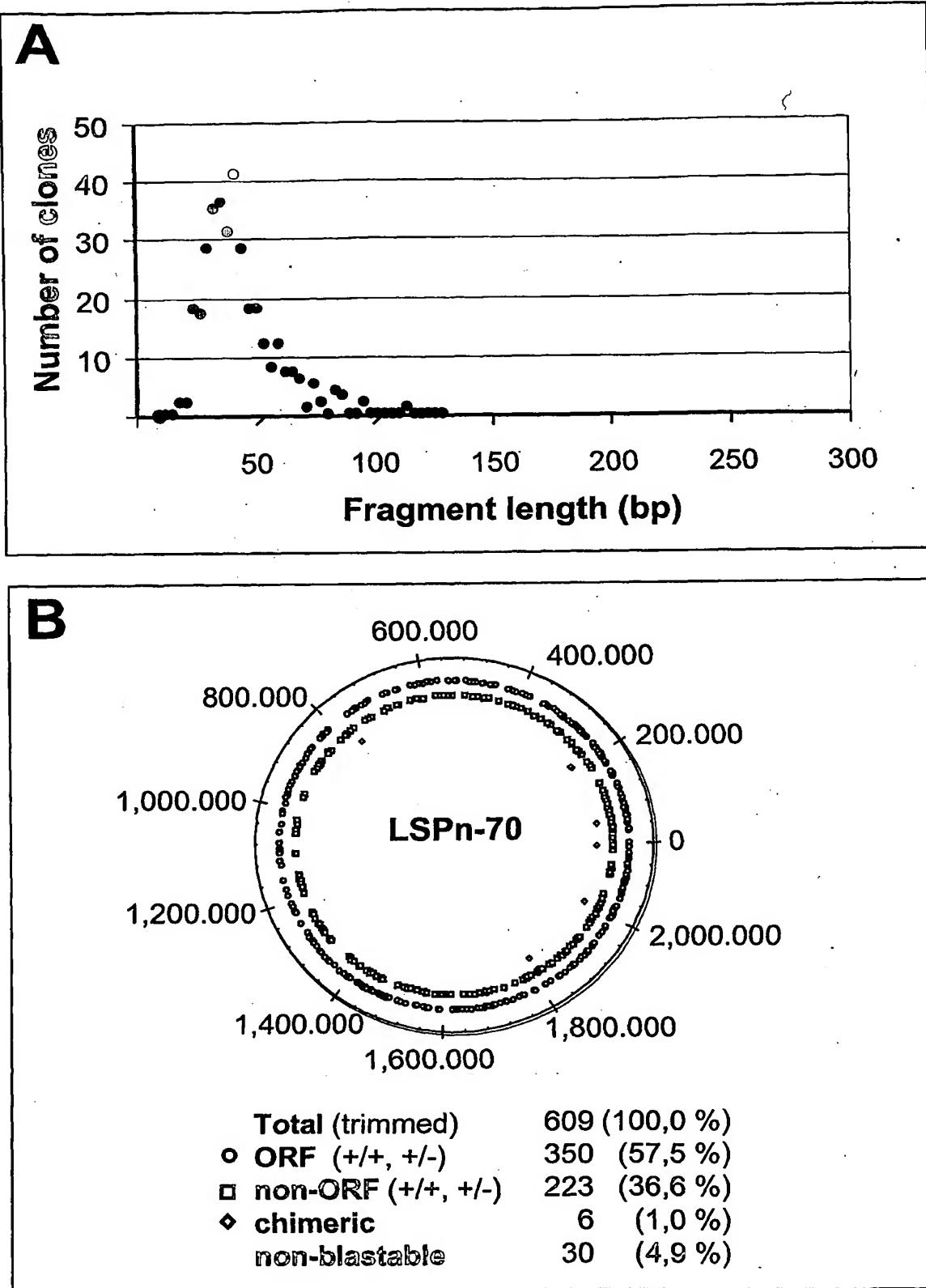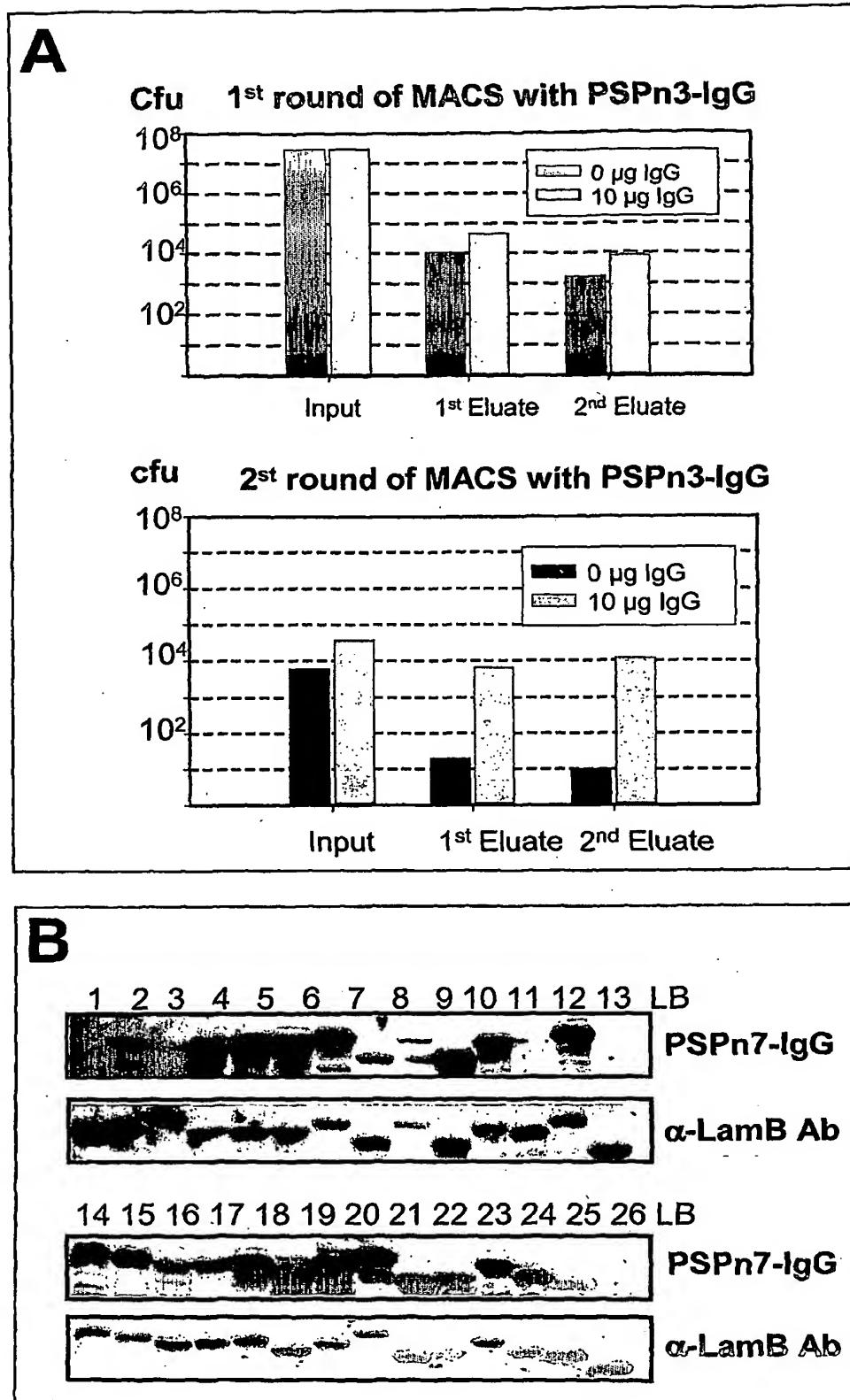
Fig. 3
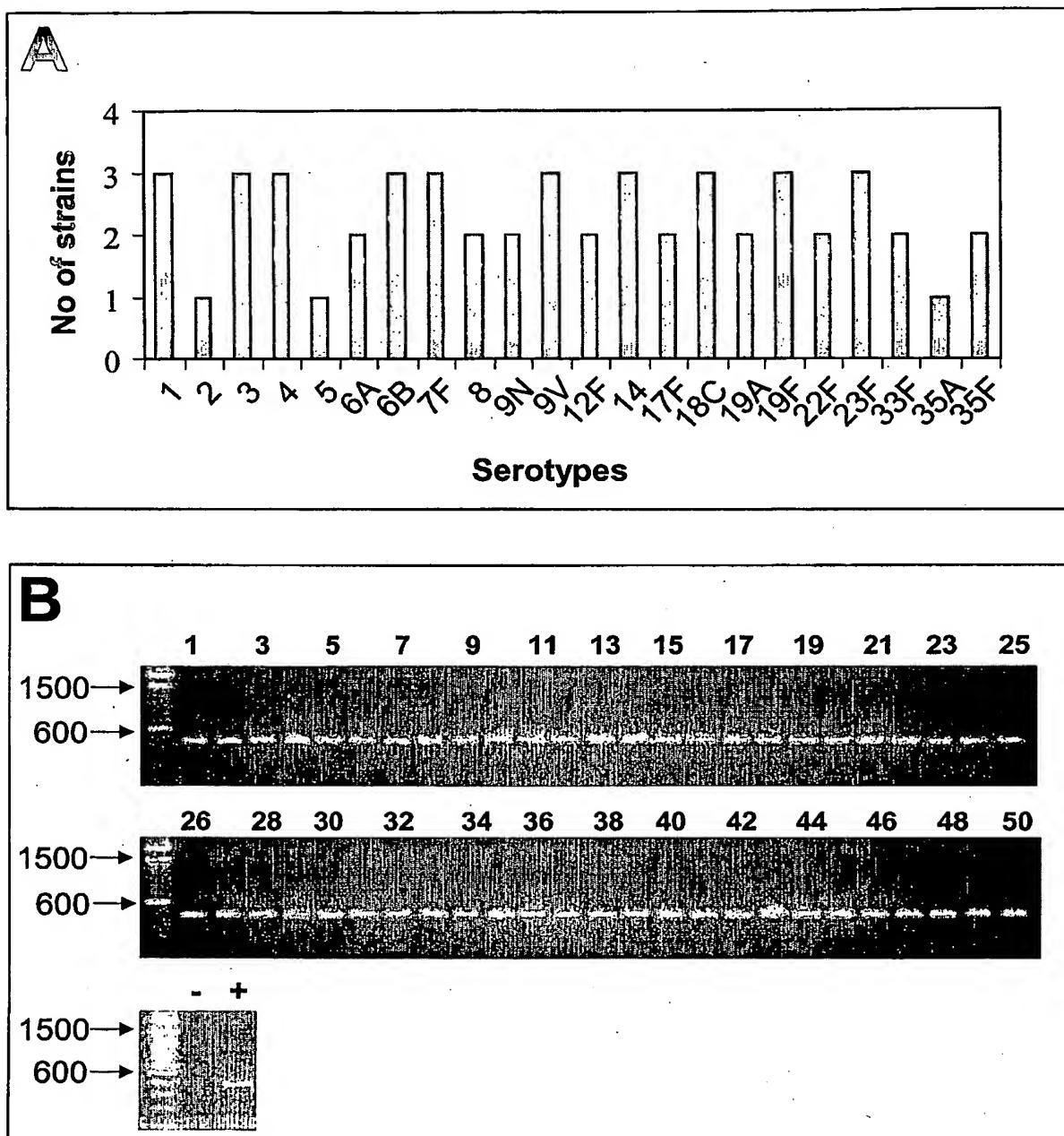
Fig. 4
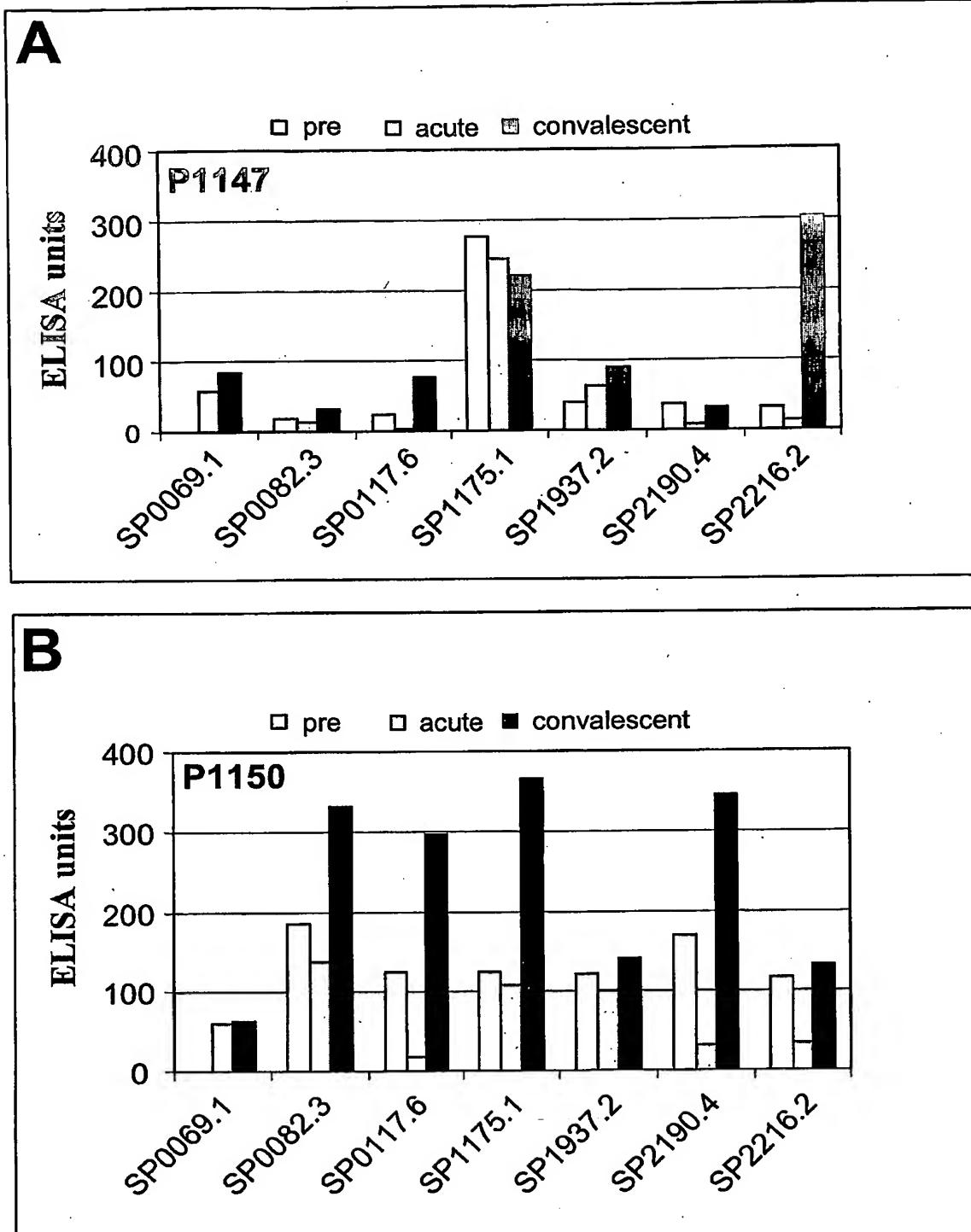
Fig. 5
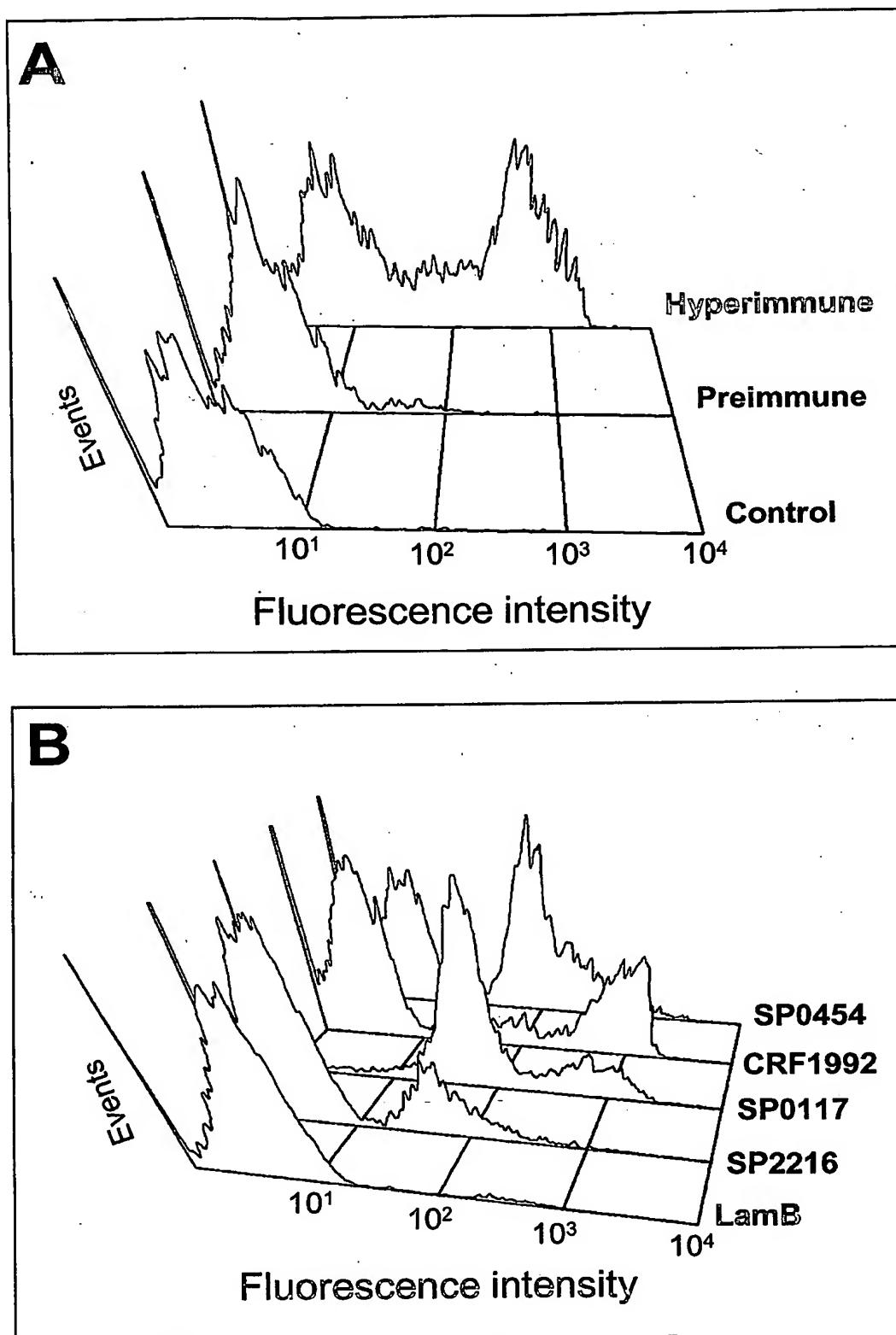
Fig. 6
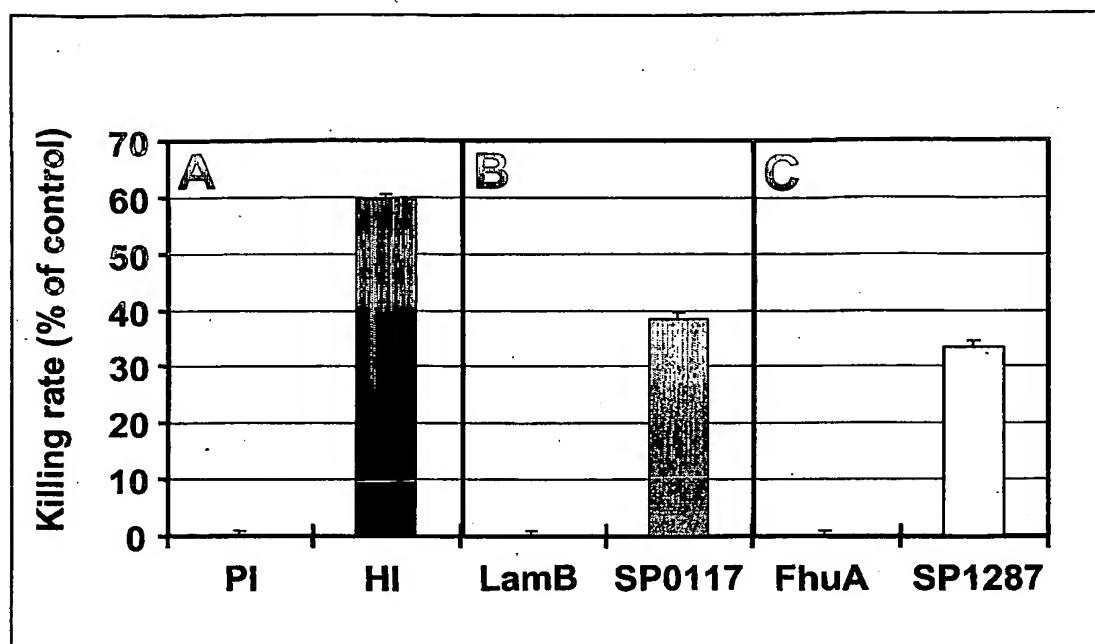
Fig. 7
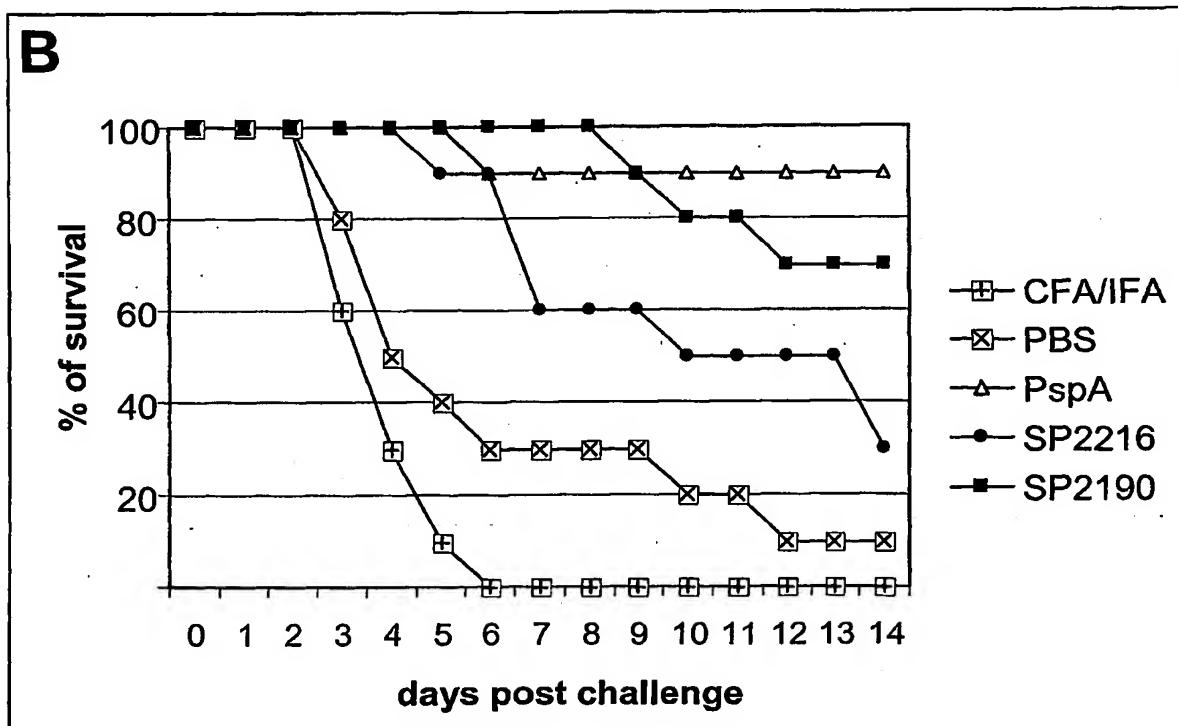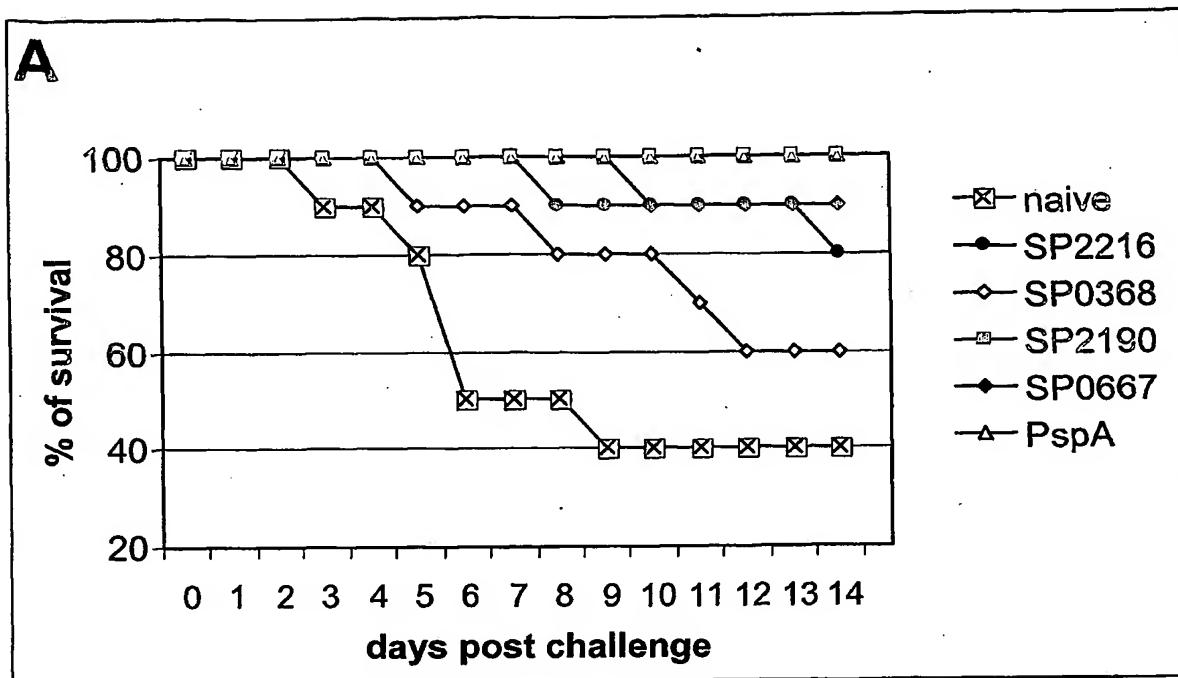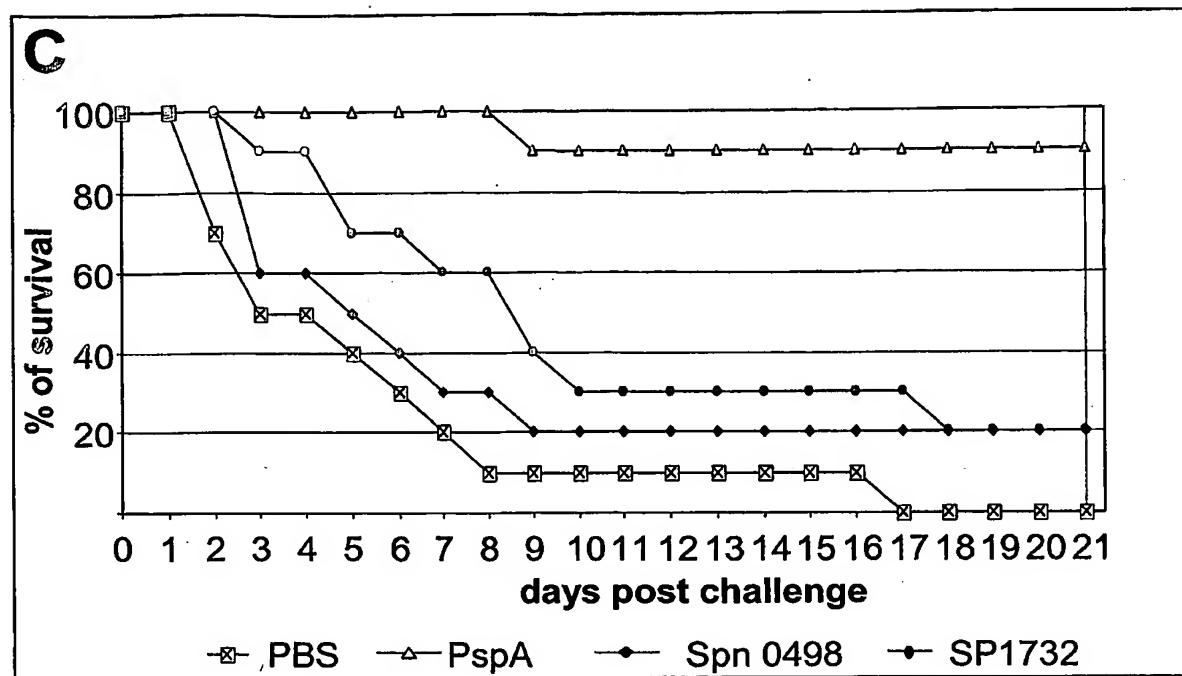
Fig. 8

Fig. 8
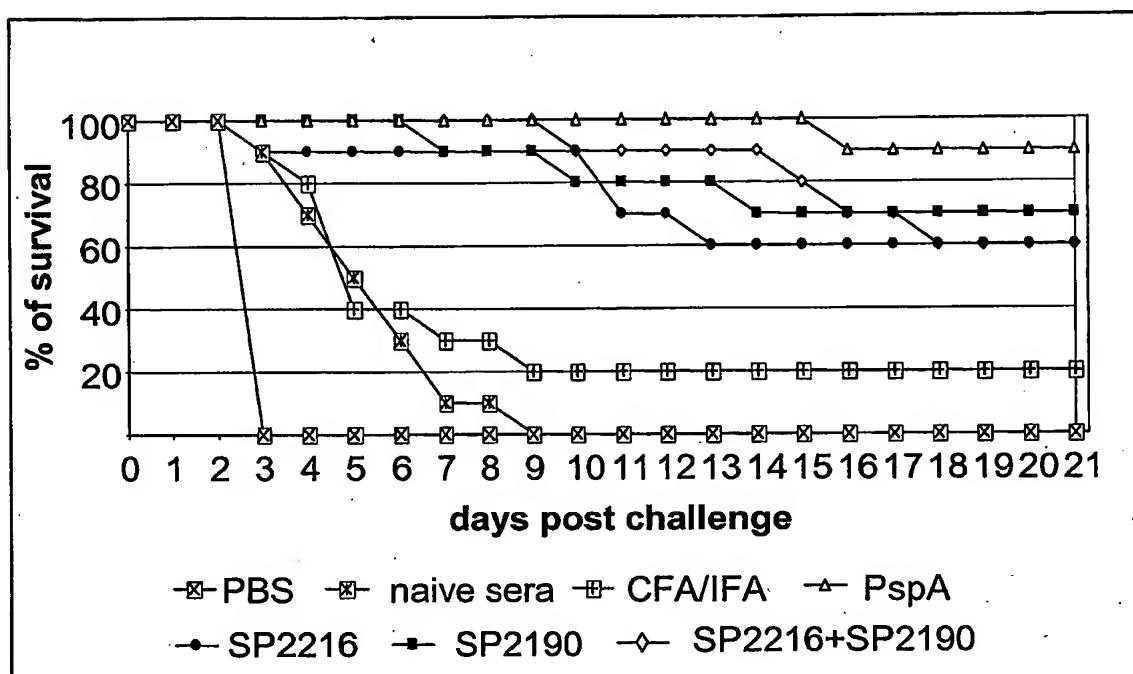
Fig. 9
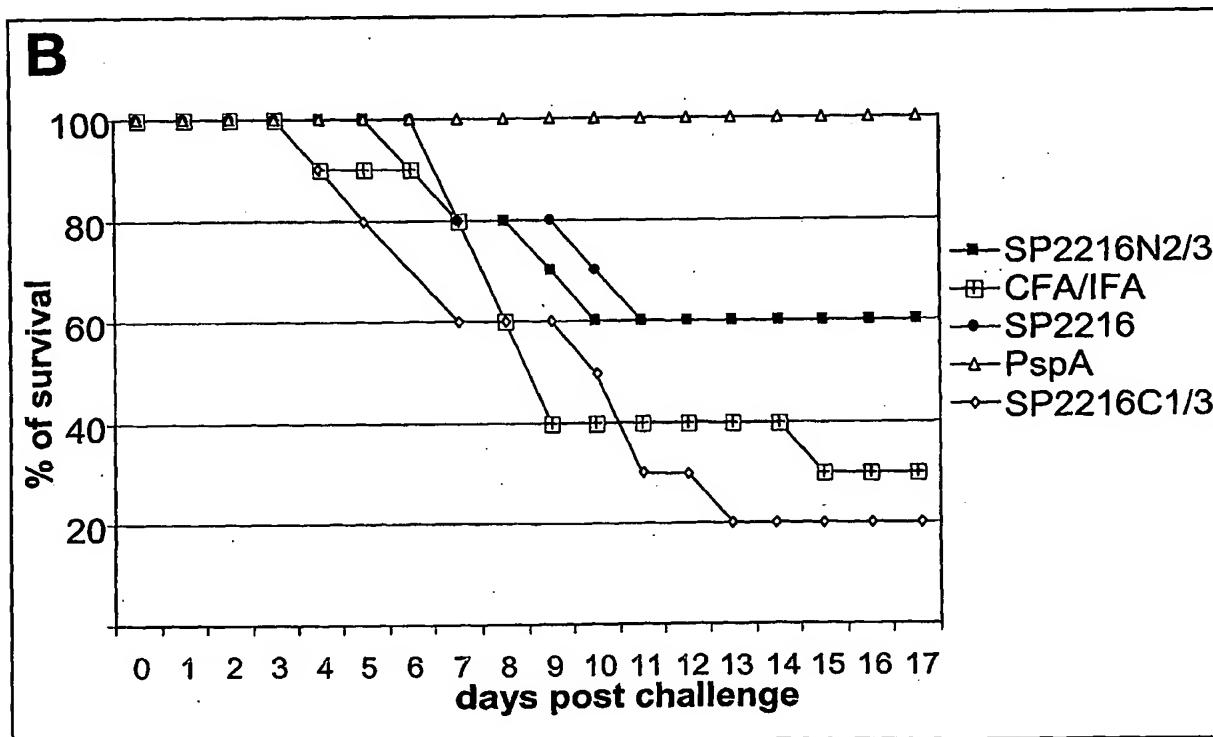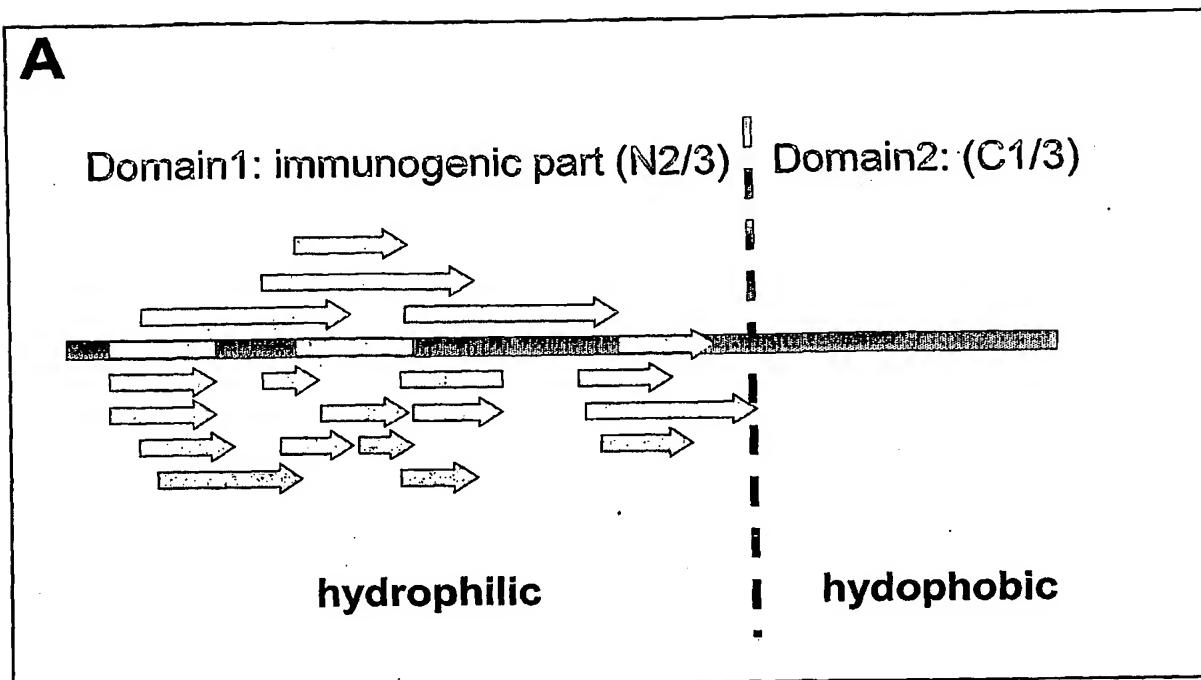
Fig. 10
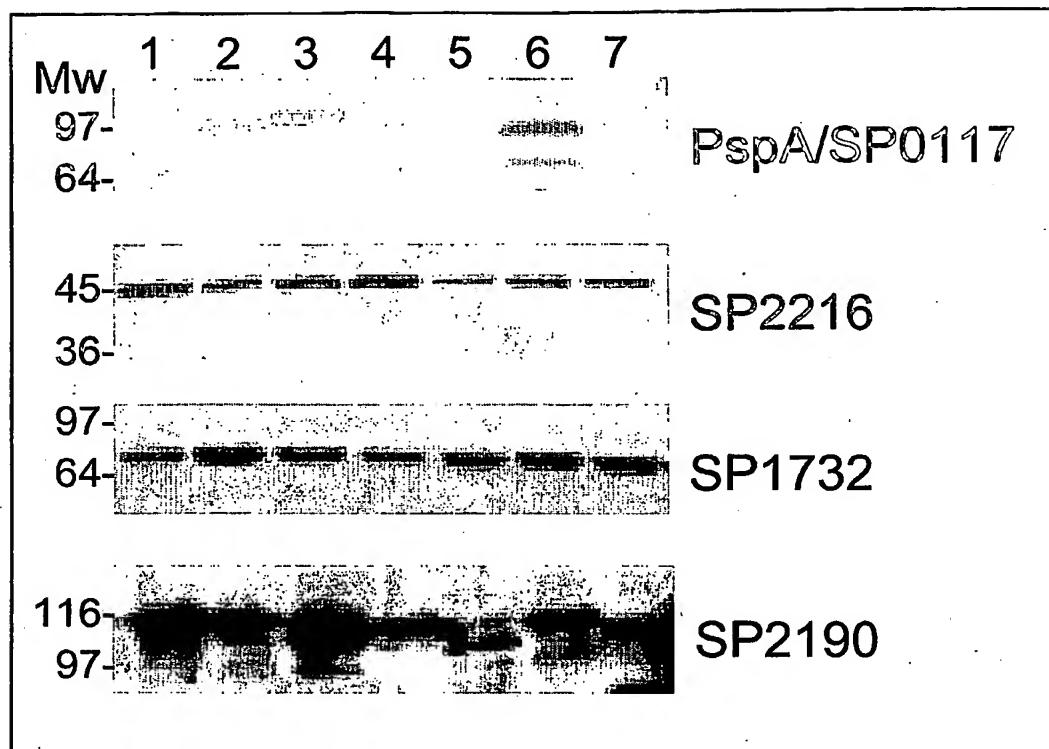
Fig. 11

```
                   1                                                                          90
TIGR4      MKKKILASLL LSTVMVSQVA VLTTAHAETT DDKIAAQDNK ISNLTAQQQE AQKQVDQIQE QVSAIQAEQS NLQAENDRLQ AESKKLEGEI
   14      MKKKILASLL LSTVMVSQVA VLTTAHAETT DDKIAAQDNK ISNLTAQQQE AQKQVDQIQE QVSAIQAEQS NLQAENDRLQ AESKKLEGEI
   19F     MKKKILASLL LSTVMVSQVA VLTTAHAETT DDKIAAQDNK ISNLTAQQQE AQKQVDQIQE QVSAIQAEQS NLQAENDRLQ AESKKLEGEI
Consensus  mKKKILASLL LSTVMVSQVA VLTTAHAETT DDKIAAQDNK ISNLTAQQQE AQKQVDQIQE QVSAIQAEQS NLQAENDRLQ AESKKLEGEI

                   91                                                                         180
TIGR4      TELSKNIVSR NQSLEKQARS AQTNGAVTSY INTIVNSKSI TEAISRVAAM SEIVSANNKM LEQQKADKKA ISEKQVANND AINTVIANQQ
   14      TELSKNIVSR NQSLEKQARS AQTNGAVTSY INTIVNSKSI TEAISRVAAM SEIVSANNKM LEQQKADKKA ISEKQVANND AINTVIANQQ
   19F     TELSKNIVSR NQSLEKQARS AQTNGAVTSY INTIVNSKSI TEAISRVAAM SEIVSANNKM LEQQKADKKA ISEKQVANND AINTVIANQQ
Consensus  TELSKNIVSR NQSLEKQARS AQTNGAVTSY INTIVNSKSI TEAISRVAAM SEIVSANNKM LEQQKADKKA ISEKQVANND AINTVIANQQ

                   181                                                                        270
TIGR4      KLADDAQALT TKQAELKAAE LSLAAEKATA EGEKASLLEQ KAAAEAEARA AAVAEAAYKE KRASQQQSVL ASANTNLTAQ VQAVSESAAA
   14      KLADDAQALT TKQAELKAAE LSLAAEKATA EGEKASLLEQ KAAAEAEARA AAVAEAAYKE KRASQQQSVL ASANTNLTAQ VQAVSESAAA
   19F     KLADDAQALT TKQAELKAAE LSLAAEKATA EGEKASLLEQ KAAAEAEARA AAVAEAAYKE KRAGQQQSVL ASANTNLTAQ VQAVSESAAA
Consensus  KLADDAQALT TKQAELKAAE LSLAAEKATA EGEKASLLEQ KAAAEAEARA AAVAEAAYKE KRAsQQQSVL ASANTNLTAQ VQAVSESAAA

                   271                                                                        360
TIGR4      PVRAKVRPTY STNASSYPIG ECTWGVKTLA PWAGDYWGNG AQWATSAAAA GFRTGSTPQV GAIACWNDGG YGHVAVVTAV ESTTRIQVSE
   14      PVRAKVRPTY STNASSYPIG ECTWGVKTLA PWAGDYWGNG AQWATSAAAA GFRTGSTPQV GAIACWNDGG YGHVAVVTAV ESTTRIQVSE
   19F     PVRAKVRPTY STNASSYPIG ECTWGVKTLA PWAGDYWGNG AQWATSAAAA GFRTGSTPQV GAIACWNDGG YGHVAVVTAV ESTTRIQVSE
Consensus  PVRAKVRPTY STNASSYPIG ECTWGVKTLA PWAGDYWGNG AQWATSAAAA GFRTGSTPQV GAIACWNDGG YGHVAVVTAV ESTTRIQVSE

                   361                  392
TIGR4      SNYAGNRTIG NHRGWFNPTT TSEGFVTYIY AD
   14      SNYAGNRTIG NHRGWFNPTT TPEGFVTYIY AD
   19F     SNYAGNRTIG NHRGWFNPTT TSEGFVTYIY AD
Consensus  SNYAGNRTIG NHRGWFNPTT TsEGFVTyiy ad
```

Fig. 12

```
                                                                              90
      1
TIGR4 MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM
4     MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM
6B    MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM
9V    MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM
18C   MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM
R6    MIQIGKIFAG RYRIVKQIGR GGMADVYLAK DLILDGEEVA VKVLRTNYQT DPIAVARFQR EARAMADLDH PHIVRITDIG EEDGQQYLAM

                                                                              180
      91
TIGR4 EYVAGLDLKR YIKEHYPLSN EEAVRIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS
4     EYVAGLDLKR YIKEHYPLSN EEAARIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS
6B    EYVAGLDLKR YIKEHYPLSN EEAVRIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS
9V    EYVAGLDLKR YIKEHYPLSN EEAVRIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS
18C   EYVAGLDLKR YIKEHYPLSN EEAVRIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS
R6    EYVAGLDLKR YIKEHYPLSN EEAVRIMGQI LLAMRLAHTR GIVHRDLKPQ NILLTPDGTA KVTDFGIAVA FAETSLTQTN SMLGSVHYLS

                                                                              270
      181
TIGR4 PEQARGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQKPLP SVIAENPSVP QALENVIIKA TAKKLTNRYR SVSEMYVDLS
4     PEQAGGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQKPLP SVIAENPSVP QALENVIIKA TAKKLTNRYR SVSEMYVDLS
6B    PEQARGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQKPLP SVIAENPSVP QALENVIIKA TAKKLTNRYR SVSEMYVDLS
9V    PEQARGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQKPLP SVIAENPSVP QALENVIIKA TAKKLTNRYR SVSEMYVDLS
18C   PEQARGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQKPLP SVIAENPSVP QALENVIIRA TAKKLTNRYR SVSEMYVDLS
R6    PEQARGSKAT VQSDIYAMGI IFYEMLTGHI PYDGDSAVTI ALQHFQNPLP SVIAENSSVP QALENVIIKA TAKKLTNRYR SVSEMYVDLS

                                                                              360
      271
TIGR4 SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTEHKSIK NPSQAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
4     SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTEHKSIK NPSQAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
6B    SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTEHKSIK NPSQAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
9V    SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTEHKSIK NPSRAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
18C   SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTGHKSIK NPSQAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
R6    SSLSYNRRNE SKLIFDETSK ADTKTLPKVS QSTLTSIPKV QAQTEHKSIK NPSQAVTEET YQPQAPKKHR FKMRYLILLA SLVIVAASLI
```

Fig. 13

```
                                                                                          450
361
TIGR4  WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD
4      WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD
6B     WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD
9V     WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD
18C    WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD
R6     WILSRTPATI AIPDVAGQTV AEAKATLKKA NFEIGEEKTE ASEKVEEGRI IRTDPGAGTG RKEGTKINLV VSSGKQSFQI SNYVGRKSSD

                                                                                          540
451
TIGR4  VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQIV LTVAKKATTI QLGNYIGRNS TEVISELKQK KVPENLIKIE
4      VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQI- ---------- ---------- ---------- ----------
6B     VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQIV LTVAKKATTI QLGNYIGRNS TEVISELKQK KVPENLIKIE
9V     VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQIV LTVAKKATTI QLGNYIGRNS TEVISELKQK KVPENLIKIE
18C    VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQII LTVAKKATTI QLGNYIGRNS TEVISELKQK KVPENLIKIE
R6     VIAELKEKKV PDNLIKIEEE ESNESEAGTV LKQSLPEGTT YDLSKATQIV LTVAKKATTI QLGNYIGRNS TEVISELKQK KVPENLIKIE

                                                                               630
541
TIGR4  EEESSESEPG TIMKQSPGAG TTYDVSKPTQ IVLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSAEGMVVEQ
4      ---------- ---------- ---------- -VLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSAEGMVVEQ
6B     EEESSESEPG TIMKQSPGAG TTYDVSKPTQ IVLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSVEGMVVEQ
9V     EEESSESEPG TIMKQSPGAG TTYDVSKPTQ IVLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSAEGMVVEQ
18C    EEESSESEPG TIMKQSPGAG TTYDVSKPTQ IVLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSAEGMVVEQ
R6     EEESSESEPG TIMKQSPGAG TTYDVSKPTQ IVLTVAKKVT SVAMPSYIGS SLEFTKNNLI QIVGIKEANI EVVEVTTAPA GSVEGMVVEQ

                              659
631
TIGR4  SPRAGEKVDL NKTRVKISIY KPKTTSATP
4      SPRAGEKVDL NKTRVKISIY KPKTTSATP
6B     SPRAGEKVDL NKTRVKISIY KPKTTSATP
9V     SPRAGEKVDL NKTRVKTSIY KPKTTSATP
18C    SPRAGEKVDL NKTRVKISIY KPKTTSATP
R6     SPRAGEKVDL NKTRVKISIY KPKTTSATP
```

Fig. 13

A

```
            1                                                                          90
TIGR4     MFASKSERKV HYSIRKFSVG VASVVVASLV MGSVVHATEN EGATQVPTSS NRANESQAEQ GEQPKKLDSE RDKARKEVEE YVKKIVGESY
R6        MFASKSERKV HYSIRKFSIG VASVAVASLV MGSVVHATEN EGSTQAATSS NMAKTEH--- ---------- RKAAKQVVDE YIEKMLRE--
4         MFASKSERKV HYSIRKFSVG VASVVVASLV MGSVVHATEN EGATQVPTSS NRANESQAEQ GEQPKKLDSE RDKARKEVEE YVKKIVGESY
9V        MFASKSERKV HYSIRKFSVG VASVAVASLV MGSVVHATEN ERTTQVPTSS NRGKPER--- ---------- RKAAEQF-DE YINKM-----
14        MFASKSERKV HYSIRKFSIG VASVAVASLF LGGVVHA-EG VRSENTPKVT SSGDE----- ---------- --VDE YIKKMLSE--
18C       MFASKSERKV HYSIRKFSIG VASVAVASLV MGSVVHATEK EVTTQVPTYS NMAKTEH--- ---------- RKAAKQVVDE YIEKMLRE--
19F       MFASKSERKV HYSIRKFSIG VASVAVASLF LGGVVHA-EG VRSENTPKVT SSGDE----- ---------- --VDE YIKKMLSE--
23F       MFASKSERKV HYSIRKFSIG VASVAVASLF MGSVVHATEK EVTTQVATSS NKANKSQ--- ---------TE HMKAAKQVDE YIKKKL----
Consensus MFASKSERKV HYSIRKFS!G VASVaVASLv $GsVVHAtE. e..t#vptss n.a.e..... r..a...v#E Y!kKml.e..
            91                                                                         180
TIGR4     AKSTKKRHTI TVALVNELNN IKNEYLNKI- VESTSESQLQ ILMMESRSKV DEAVSKFEKD SSSSSSSDSS TKPEASDTAK PNKPTEPGEK
R6        IQLDRRKHTQ NVALNIKLSA IKTKYLREL- -NVLEEKSKD ELPSEIKAKL DAAFEKFKKD T--------- ---------- ----LKPGEK
4         AKSTKKRHTI TVALVNELNN IKNEYLNKI- VESTSESQLQ ILMMESRSKV DEAVSKFEKD SPSSSSSDSS TKPE
9V        IQLDKRKHTQ NLAFNIQLSR IKTEYLNGL- ----KEKSEA ELPSKIKAEL DAAFKQFKKD TLPTEPE--- ---------- --KKVAEAKK
14        IQLDKRKHTH NFALNLKLSR IKTEYLYKLK VNVLEEKSKA ELTSKTKKEV DAAFEKFKKD T--------- ---------- ----LKLGEK
18C       IQLDRRKHTQ NFAFNMKLSA IKTEYLIYGL- ----KEKSEA ELPSS-EAEL PSEVKA-KLD AAFEQFK--- ---------- -KDTLKLGEK
19F       IQLDRRKHTH NFALNLKLSR IKTEYLYKLK VNVLEEKSKA ELTSKTKKEV DAAFEKFKKD T--------- ---------- ----LKLGEK
23F       -QLDRRKHTQ NVGLLITKLGV IKTEYLHGL- -SVSKKKSEA ELPSEIKAKL DAAFBQFKKD TLP------- ---------- ----TEPGKK
Consensus iqldkrkHT. n.aln.kLs. IKtEYl.kl. v....eks.a eL.s..k.ev daaf.kfkkD t......... ..........gek
            181                                                                        270
TIGR4     VAEAKKKVEE AEKKAKDQKE EDRRNYPTIT YKTLELEIAE SDVEVKKAEL ELVKVKANEP RDEQKIKQAE AEVESKQAEA TRLKKIKTDR
R6        VAEAKKKVEE AKKKAEDQKE EDRRNYPTNT YKTLELBIAE FDVKVKEAEL ELVKEEAKES RNEGTIKQAK BKVESKKAEA TRLENIKTDR
9V        VEEAKKKVAE AKKKAKAQKE EDHRNYPTIT YKTLDLEIAE FDVKVKEAEL ELVKKEADES RNEGTINQAK AKVESEKAEA TRLKKIKTDR
14        VAEAQKKVEE AKKKAKDQKE EDHRNYPTNT YKTLELEIAE SDVKVKKAEL ELLKEEA-KT RNEDTINQAK AKVKSEQAEA TRLKKIKTDR
18C       VAEAEKKVAE AKKKAKAQKE EDRRNYPTIT YKTLDLEIAE SDVEVKKAEL ELLKEEA-KT RNKDTIKQAK ARVESKKAEA TKLEBIKTDR
19F       VAEAQKKVEE AKKKAKDQKE EDHRNYPTNT YKTLELEIAE SDVKVKKAEL ELLKEEA-KT RNEDTINQAK AKVKSEQAEA TRLKKIKTDR
23F       VAEAEKKVEE AKKKAEDQKE KDLRNYPTNT YKTLELDIAE SDVEVKKAEL ELVKGSYRNL ETRKKLIKQS EKLRIKKLML Q
Consensus vaea.kkvee akkkakdqke ed.rnypt.t yktlelelae sdv.vk.ael el.k.ea... rne.tl.qak akv.s..aea trl..iktdr
            271                                                                        360
TIGR4     EEAEEE-AKR RADAK----- ---EQGKPK GRAKRGVPGE LATPDKKEND AKSSDSSVGE ETLPSPSLKP EKKVAEAEKK VEEAKKKAED
R6        KKAEEE-AKR KADAKLKEAN VATSDQGKPK GRAKRGVPGE LATPDKKEND AKSSDSSVGE ETLPSSSLKS GKKVAEAEKK VEEAEKKAKD
9V        EKAEEEEAKR RADAKEQDES
14        EQAEAT
18C       KKAEEEA
19F       EQAEATRLEN IKTDREK--- ---AEEAKRK AE
Consensus ..ae......
```

Fig. 14

B

```
                                                                            450
        361
TIGR4   QKEEDRRNYP TNTYKTLELE IAESDVEVKK KEPRNEEKVK QAKAEVESKK AEATRLEKIK TDRKKAEEEA KRKAAEEDKV
R6      QKEEDRRNYP TNTYKTLDLE IAESDVKVKE KEPRDEEKIK QAKAKVESKK AEATRLENIK TDRKKAEEEA KRKAAEEDKV
4                          AESDVEVKK  KEPRNEEKVK QAKAEVESKK AEATRLEKIK TDRKKAEEEA KRRAAEEDKV
6B             LK LLSPMWKLKK RSLTS-KEEA KKPLNEGTIR QAKAEVESKK AEATRLEKIK TDRKKAEEEA KRKAAEEDKV
9V                         IAESDVKVKE KESRNEEKIK QAKAKVESKK AEATRLEKIK TDRKKAEEEA KRKAAEEDKV
14                         IAESDVKVKE KEPRDEEKIK QAKAEVESKQ AEATRLEKIK TDRKKAEEEA KRKAAEEDKV
18C                        IAESDVKVKE KESRNEEKVK QAKAKVESKK AEATRLEKIK TDRKKAEEE-A KRKAAEEDKV
19F            LE IAESDVKVKE KESRNEEKVK QAKAKVESKQ AEATRLEKIK TDRKKAEEEA KRKAAEEDKV
23F

Consensus  ..........  iaesdvkvke keprneek.k qaka.veskk aeatrlekik tdrkkaeeea krkaaeedkv

                                                                            540
        451
TIGR4   KEKPAEQPQP APAPKAEKPA ------PAP KPENPAEQPK AEKPA----- ---------- --DQQAEED  YARRSEEEYN RLTQQQPPKT
R6      KEKPAEQPQP APATQPEKPA ------P-- KPEKPAEQPK AEKTD----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKT
4       KEKPAEQPQP APAPKAEKP- -------APAP KPENPAEQPK AEKPA----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKT
6B      KEKPAEQPQP APAPQPEKPT EEPENPVPAP KPEKPAEQPK PEKPAEQPKP EKPAEQPKAB KTDDQQAEED YARRSEEEYN RLTQQQPQKP
9V      KEKPAEQPQP APAPKPENPA EEPENPVPAP KPENPAEQPK AEKPA----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKP
14      KEKPAEQPQP APAPQPEKPT PKPEKPAPAP KPENPAEQPK AEKPA----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKT
18C     KEKPAEQPQP APAPQPEKPT EEPENPAPAP KPEKPAEQPK AEKPA----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKT
19F     KEKPAEQPQP APAPQPEKP- -------APAP KPENPAEQPK AEKPA----- ---------- ---DQQAEED YARRSEEEYN RLTQQQPPKT
23F                                                            VQNGMW VYFYNTDGSMA TGWLQNNGSW

Consensus kekpaeqpqp apapqpekp. ........pap kpenpaeqpk aekpa..... ......... ...dqQaeed Yarrs#eeyn rltq#ppkt

                                                                            630
        541
TIGR4   EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQNNGSW YYLNSNGAMA TGWLQNNGSW YYLNANGSMA
R6      EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQNNGSW YYLNANGSW  TGWLQNNGSW YYLNANGAMA
4       EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQNNGSW YYLNSNGAMA TGWLQNNGSW YYLNANGSMA
6B      EQPAP--APK IGWKQENGMW YFYNTDGSMA TGWLQ----- ---------- --------- ---------
9V      EQPAP--APK IGWKQENGMW YFYNTDGSMA TGWLQNNGSW YYLNANGAMA TGWLQNNGSW YYLNANGDMA
14      EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQYNGSW YYLNANGDMA TGWLQNNGSW YYLNANG---
18C     EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQNNGSW YYLNANGDMA TGWLQNNGSW YYLNANGDMA
19F     EKPAQPSTPK TGWKQENGMW YFYNTDGSMA TGWLQYNGSW YYLNANGDMA TGWLQNNGSW YYLNANG---
23F     YYLNSNGAMA TGWLQNNGSW YYLNSNGAMA TGWFQYNGSW YYLNANGDMA

Consensus ekpaqpstpk tgWKQ#NGmW Y%yNt#GsMA TGWLQ.ngsw yyln.ng.ma tgwlqnngsw yylnang.ma
```

Fig. 14

```
          631                                                                        720
TIGR4     TGWLQYNGSW YYLN------ ---------- ----ANGSMA TGWLQYNGSW YYLNANGDMA TGWVKDGDTW
R6        TGWLQYNGSW YYLNSNGAMA TGWLQYNGSW YYLNANGDMA TGWLQYNGSW YYLNANGDMA TGWVKDGDTW
4         TGWLQYNGSW YYLN------ ---------- ----ANGSMA TGWLQYNGSW YYLNANGDMA TGWVKDGDTW
6B        ---------- ---------- ---------- ---------- -----YNGSW YYLNANGSMA TGWVKDGDTW
9V        TGWLQYNGSW YYLN------ ---------- ----ANGDMA TGWLQYNGSW YYLNANGDMA TGWVKDGDTW
14        ---------- ---------- ---------- -------DMA TGWLQNNGSW YYLNANGDMA TGWVKDGDTW
18C       TGWLQYNGSW YYLN------ ---------- ----ANGDMA TGWLQNNGSW YYLNANGDMA TGWVKDGDTW
19F       ---------- ---------- ---------- -------DMA TGWLQNNGSW YYLNANGDMA TGWVKDGDTW
23F       TGWFQYNGSW YYLN------ ---------- ----ANGDMA TGWLQYNGSW YYLNSNGAMV TDWVKDGDTW
Consensus tgw.qyngsw yyln ................ngdma ......... tgwlqyNGSW YYLNANGdMA TgWVKDGDTW

          721                                              769
TIGR4     YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
R6        YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
4         YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
6B        YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
9V        YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
14        YYLEASGAMK ASQWFKASDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
18C       YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
19F       YYLEASGAMK ASQWFKASDK WYYVNGSGAL AVNTTVDGYG VNANGEWVN
23F       YYLEASGAMK ASQWFKVSDK WYYVNGSGAL AVNTTVDSYR VNANGEWVN
Consensus YYLEASGAMk asqwfkvsdk wyyvngsgal avnttvdgyg vnangewvn
```

Fig. 14

SEQUENCE LISTING

```
SeqID 1
atgtctaaaa atattgtaca attgaataat tcttttattc aaaatgaata ccaacgtcgt      60
cgctacctga tgaaagaacg acaaaaacgg aatcgtttta tgggaggggt attgattttg     120
attatgctat tatttatctt gccaactttt aatttagcgc agagttatca gcaattactc     180
caaagacgtc agcaattagc agacttgcaa actcagtatc aaactttgag tgatgaaaag     240
gataaggaga cagcatttgc taccaagttg aaagatgaag attatgctgc taaatataca     300
cgagcgaagt actattattc taagtcgagg gaaaaagttt atacgattcc tgacttgctt     360
caaagg                                                                366


SeqID 2
atggataaga aaaaattatt attgattgat gggtcttctg tagcttttcg ggcgtttttt      60
gcgctgtatc agcagttgga ccgtttttaag aatgtggctg gtttgcatac caatgcgatt     120
tatggttttc agttgatgtt gagtcattta ttggagcggg ttgagccgag tcatattttg     180
gtggcttttg atgcgggaaa gacgacccttc cggacagaga tgtatgcgga ctataagggt     240
ggtcgggcca agactcctga tgagtttcgt gagcaatttc ctttcattcg tgagttgctg     300
gatcatatgg ggattcgtca ctatgatctg gctcagtatg aggcggatga catcattggg     360
acgctggata agctagcaga gcaggatggt tttgatatta ctattgtcag tgggggacaag     420
gatttgattc agctgacgga tgagcatacg gtggttgaaa tttccaagaa aggtgtggct     480
gagtttgagg cctttacgcc agattacctc atggaagaaa tgggcctcac accagctcag     540
tttatcgatc tcaaggcgct catgggtgat aagtcggata atatccctgg ggtgaccaaa     600
gtcggtgaaa agacgggtat taagctcttg ctggagcatg ttcgcttga ggggatttat     660
gaaaatattg atggaatgaa gacttctaag atgaaggaaa atctcatcaa tgacaaggaa     720
caggcctttt tgtctaaaac actagcgacc attgatacca aggcaccgat tgcgattggt     780
ttagaggact tggtctatag tggtccagat gttgaaaatc ttgggaaatt ctacgatgag     840
atgggcttca aacagctaaa gcaggcttta aatgtgtcgt cagctgatgt gtctgagagt     900
ttggatttta ctattgttga ccaaatcagt caagatatgc tgagtgaaga gtctatcttc     960
cactttgagc tttttggtga gaattaccat acggataatt ggtttggatt tgtctggtct    1020
tgtggggata agctctatgc cacagacaag cttgagctgt tgcaagaccc gattttcaag    1080
gatttcttag aaaaaacatc tctgagagtt tatgacttta agaaggttaa agttcttttg    1140
caacgttttg gtgtggattt gcaggcgcct gctttttgaca tccgtttggc taaatacctc    1200
ctttcgactg tggaggacaa tgaaattgcg accatcgcta gtctttatgg tcagacttac    1260
ttggttgatg atgaaacttt ctacggtaag ggtgttaaaa aggccattcc tgaacgtgag    1320
aaattcttgg aacacttagc ttgtaaactt gctgttttgg tagaaacaga gcctatttta    1380
cttgaaaaac tcagcgaaaa tgggcaatta gagcttcttt atgatatgga gcaacctctg    1440
gctttttgtcc ttgccaagat ggaaattgct gggattatgg tcaagaaaga gaccttgctt    1500
gagatgcagg ctgaaaatga gcttgtcatt gaaaaactga ctcaagagat ttacgagctg    1560
gctggtgagg agtttaatgt caactcgcct aagcagttgg gcgtgcttct cttttgagaaa    1620
ttgggacttc ctctagaata cactaagaaa accaagacag gttattcgac agcagtggat    1680
gttttagagc gtctcgctcc tattgctccg attgttaaga aaatcctgga ttaccgtcaa    1740
attgctaaga ttcaatctac ttatgtaatt ggcttgcagg actggatttt ggctgatgga    1800
aagattcata ctcgctatgt gcaggatttg acccagaccg ggcgtttgtc tagtgtggat    1860
ccaaacttgc aaaatattcc tgcccgattg gaacaggggc gcttgattcg gaaggctttt    1920
gtgccagagt gggaggatag tgtgctactc agctctgact attcacagat tgaattgcgc    1980
gttttggcgc atatttctaa ggatgagcac ttgattaagg ccttccaaga ggggcagat    2040
atccatactt cgacagccat gcgggtcttt ggcattgagc gtcctgatga tgtgactgca    2100
aacgaccgtc gcaatgccaa ggcagttaac tttggagtgg tttatgggat ttcagacttt    2160
ggcttgtcta taaatttggg aattagtcgt aaggaagcca agcctacat tgatacctac    2220
tttgaacgtt ttccaggtat taaaaactac atggatgaag tggtgcggga ggcgcgtgat    2280
aagggctatg tagagaccct ctttaagcgt cgccgtgagt tgccagatat caattcgcgc    2340
aacttcaata ttcgtggttt tgcggagcga actgctatca actcacctat ccagggttcg    2400
gcagcagata ttctcaagat tgccatgatt cagctggata aagccttggt tgcaggtggt    2460
tatcagacta gatgctgtt acaagtgcac gatgaaatcg tccttgaagt gcctaaatct    2520
gaattggtag atgaaaaa attggtgaaa caaaccatgg aagaagccat tcaactcagt    2580
gttcctctta tcgcagatga gaatgaaggg caacctggt acgaggctaa a            2631


SeqID 3
atggggatgg cagcttttaa aaatcctaac aatcaataca aagctattac aattgctcaa      60
actctaggtg atgatgcttc ttcagaggaa ttggctggta gatatggttc tgctgttcag     120
tgtacagaag tgactgcctc aaaccttttca acagttaaaa ctaaagctac ggttgtagaa     180
aaaccactga aagattttag agcgtctacg tctgatcagt ctggttgggt ggaatctaat     240
ggtaaatggt atttctatga gtctggtgat gtgaagacag gttgggtgaa aacagatggt     300
aaatggtact atttgaatga cttaggtgtc atgcagactg gatttgtaaa attttctggt     360
agctggtatt acttgagcaa ttcaggtgct atgtttacga gctgggaac agatggtagc     420
agatggttct actttgacgg ctcaggagct atgaagacag gctggtacaa ggaaaatggc     480
acttggtatt accttgacga agcaggtatc atgaagacag gttggtttaa agtcggacca     540
cactggtact atgcctacgg ttcaggagct tttggctgtga gcacaacaac accagatggt     600
```

```
taccgtgtaa atggtaatgg tgaatgggta aac                              633

SeqID 4
atgagccgaa aaagcattgg tgagaaacgc catagtttct cgatgagaaa gttgtcagtg    60
ggattggtat cagttactgt atctagtttc tttttgatga gtcaagggat tcaatcggta   120
tcggccgata atatggaaag tccaattcat tataagtata tgaccgaggg taaattgaca   180
gacgaggaaa aatccttgct ggtagaggcc cttccacaac tggctgaaga atcagatgat   240
acttattact tggtttatag atctcaacag tttttaccga atacaggttt taacccaact   300
gttggtactt tccttttttac tgcaggattg agcttgttag ttttattggt ttctaaaagg   360
gaaaatggaa agaaacgact tgttcatttt ctgctgttga ctagcatggg agttcaattg   420
ttgccggcca gtgcttttgg gttgaccagc cagattttat ctgcctataa tagtcagctt   480
tctatcggag tcggggaaca tttaccagag cctctgaaaa tcgaaggtta tcaatatatt   540
ggttatatca aaactaagaa acaggataat acagagcttt caaggacagt tgatgggaaa   600
tactctgctc aaagagatag tcaaccaaac tctacaaaaa catcagatgt agttcattca   660
gctgatttag aatggaacca aggacagggg aaggttagtt tacaaggtga agcatcaggg   720
gatgatggac tttcagaaaa atcttctata gcagcagaca atctatcttc taatgattca   780
ttcgcaagtc aagttgagca gaatccggat cacaaaggag aatctgtagt tcgaccaaca   840
gtgccagaac aaggaaatcc tgtgtctgct acaacggtgc agagtgcgga gaggaagta    900
ttggcgacga caaatgatcg accagagtat aaacttccat tggaaaccaa aggcacgcaa   960
gaaccggtc atgagggtga agccgcagtc cgtgaagact taccagtcta cactaagcca  1020
ctagaaacca aaggtacaca aggacccgga catgaaggtg aagctgcagt tcgcgaggaa  1080
gaaccagctt acacagaacc gttagcaacg aaaggcacgc aagagccagg tcatgagggc  1140
aaagctacag tccgcgaaga gactctagag tacacggaac cggtagcgac aaaaaggcaca  1200
caagaacccg aacatgaggg cgaagcggca gtagaagaag aacttccggc tttagaggtc  1260
actacacgaa atagaacgga aatccagaat attccttata caacagaaga aattcaggat  1320
ccaacacttc tgaaaaatcg tcgtaagatt gaacgacaag ggcaagcagg gacacgtaca  1380
attcaatatg aagactacat cgtaaatggt aatgtcgtag aaactaaaga agtgtcacga  1440
actgaagtag ctccggtcaa cgaagtcgtt aaagtaggaa cacttgtgaa agttaaacct  1500
acagtagaaa ttacaaactt aacaaaagtt gagaacaaaa aatctataac tgtaagttat  1560
aacttaatag acactacctc agcatatgtt tctgcaaaaa cgcaagtttt ccatggagac  1620
aagctagtta aagaggtgga tatagaaaat cctgccaaag agcaagtaat atcaggttta  1680
gattactaca caccgtatac agttaaaaca cacctaactt ataatttggg tgaaaataat  1740
gaggaaaata ctgaaacatc aactcaagat ttccaattag agtataagaa aatagagatt  1800
aaagatattg attcagtaga attatacggt aaagaaaatg atcgttatcg tagatatta  1860
agtctaagtg aagcgccgac tgatacggct aaatactttg taaaagtgaa atcagatcgc  1920
ttcaaagaaa tgtacctacc tgtaaaatct attacagaaa atacggatgg aacgtataaa  1980
gtgacggtag ccgttgatca acttgtcgaa gaaggtacag acggttacaa agatgattac  2040
acatttactg tagctaaatc taaagcagag caaccaggag tttacacatc ctttaaacag  2100
ctggtaacag ccatgcaaag caatctgtct ggtgtctata cattggcttc agatatgacc  2160
gcagatgagg tgagcttagg cgataagcag acaagttatc tcacaggtgc atttacaggg  2220
agcttgatcg gttctgatgg aacaaaatcg tatgccattt atgatttgaa gaaaccatta  2280
tttgatacat taaatggtgc tacagttaga gatttggata ttaaaactgt ttctgctgat  2340
agtaaagaaa atgtcgcagc gctggcgaag gcagcgaata gcgcgaatat taataatgtt  2400
gcagtagaag gaaaaatctc aggtgcgaaa tctgttgcgg gattagtagc gagcgcaaca  2460
aatacagtga tagaaaacag ctcgtttaca gggaaactta tcgcaaatca ccaggacagt  2520
aataaaaatg atactggagg aatagtaggt aatataacag gaaatagttc gagagttaat  2580
aaagttaggg tagatgcctt aatctctact aatgcacgca ataataacca aacagctgga  2640
gggatagtag gtagattaga aaatggtgca ttgatatcta attcggttgc tactgagaa   2700
atacgaaatg gtcaaggata ttctagagtc ggaggaatag taggatctac gtggcaaaac   2760
ggtcgagtaa ataatgttgt gagtaacgta gatgttggag atggttatgt tatcaccggt   2820
gatcaatacg cagcagcaga tgtgaaaaat gcaagtacat cagttgataa tagaaaagca   2880
gacagattcg ctacaaaatt atcaaaagac caaatagacg cgaaagttgc tgattatgga   2940
atcacagtaa ctcttgatga tactgggcaa gatttaaaac gtaatctaag agaagttgat   3000
tatacaagac taaataaagc agaagctgaa agaaaagtag cttatagcaa catagaaaaa   3060
ctgatgccat tctacaataa agacctagta gttcactatg gtaacaaagt agcgacaaca   3120
gataaacttt acactacaga attgttagat gttgtgccga tgaaagatga tgaagtagta   3180
acggatatta ataataagaa aaattcaata aataaagtta tgttacattt caaagataat   3240
acagtagaat acctagatgt aacattcaaa gaaaacttca taaacagtca agtaatcgaa   3300
tacaatgtta caggaaaaga atatatattc acaccagaag catttgtttc agactataca   3360
gcgataacga ataacgtact aagcgacttg caaaatgtaa cacttaactc agaagctact   3420
aaaaaagtac taggagcagc gaatgatgca gccttagata acctatactt agatagacaa   3480
tttgaagaag ttaaagctaa tatagcagaa cacctaagaa aagtattagc gatggataaa   3540
tcaatcaata ctacaggaga cggtgtagtt gaatacgtaa gtgagaaaat caaaaataac   3600
aaagaagcat ttatgctagg tcttacttat atgaaccgtt ggtacgatat taattatggt   3660
aaaatgaata caaaagattt atctacgtac aagtttgact ttaacggaaa taatgagact   3720
tcaacgttgg atactattgt cgcattagga aatagtggac tagataacct gagagcttca   3780
aatactgtag gtttatatgc gaataaactt gcatcggtaa aaggagaaga ttcagtcttt   3840
gacttcgtag aagcgtatag aaaactgttc ttaccaaaca aaacaaataa cgagtggttt   3900
```

```
aaagaaaata caaaggcata tatagtcgaa atgaagtctg atattgcaga agtacgagáa   3960
aaacaagaat caccaacagc cgatagaaaa tattcattag gagtttacga tagaatatca   4020
gcaccaagtt gggggcataa gagtatgtta ttaccactac taactttacc tgaagaatct   4080
gtgtatattt catcgaatat gtctacactt gcattcggtt cgtatgaaag atatcgtgat   4140
agtgtggatg gagttattct ttcaggagat gctttacgaa cttatgtaag aaatagagtt   4200
gatatagcag cgaaaaggca tagagaccat tatgatattt ggtacaatct tcttgacagt   4260
gcttcaaaag aaaaactttt ccgttctgtg atagtttatg atggattcaa tgtaaaagat   4320
gagacaggaa gaacttattg ggcaaggtta acggataaaa acatcggctc tattaaagaa   4380
ttcttcggac ctgttgggaa atggtatgag tataatagta gtgcaggagc gtatgcgaat  ·4440
ggaagtttaa cgcactttgt gttagataga ttattagatg cttatggaac gtcggtttat   4500
actcatgaaa tggttcataa ttctgattct gcaatctact ttgaaggaaa tggtagacgt   4560
gaaggattgg gagcggagtt atacgcactt ggtttactgc aatctgtaga tagtgtaaat   4620
tctcatattt tagctttaaa tacgttatat aaagcagaaa aagatgattt gaatagattg   4680
catacatata atccggtgga acgtttcgat tcggatgagg cgcttcaaag ttatatgcat   4740
ggatcatatg atgtaatgta tacacttgat gcgatggaag caaaagcgat attagctcaa   4800
aataatgatg ttaagaaaaa atggtttaga aaaatagaaa attattacgt tcgtgatact   4860
agacataata aagatacaca tgcaggaaat aaagtccgtc cattaacaga tgaagaagta   4920
gctaacttaa catcgttaaa ctcattcatc gacaacgaca tcataaatag acgtagctat   4980
gatgatagta gagaatataa acgaaatggc tactatacta taagtatgtt ctctcctgta   5040
tacgcagcgc taagcaattc gaaaggtgct cctggagata ttatgtttag aaaaatagct   5100
tatgaattac ttgcggaaaa aggttatcac aaaggattcc taccttatgt ttctaatcag   5160
tacggagcag aagcatttgc cagcggaagc aaaacattct catcatggca tggaagagat   5220
gttgctttag tgacagatga tttagtattt aagaaagtat tcaatggtga gtactcatca   5280
tgggctgatt tcaaaaaagc aatgtttaaa caacgtatag ataaacaaga taatctgaaa   5340
ccaataacaa ttcaatacga attaggtaat cctaatagta caaaagaagt aactataaca   5400
acggctgcac aaatgcaaca attaattaat gaagcggctg cgaaagatat tactaatata   5460
gatcgtgcaa cgagtcatac cccagcaagt tgggtgcatt tattaaaaca aaaaatctat   5520
aatgcatatc ttcgcactac agatgacttt agaaattcta tatataaa            5568
```

SeqID 5
```
atgaaattca atccaaatca aagatatact cgttggtcta ttcgccgtct cagtgtcggt    60
gttgcctcag ttgttgtggc tagtggcttc tttgtcctag ttggtcagcc aagttctgta   120
cgtgccgatg ggctcaatcc aaccccaggt caagtcttac ctgaagagac atcgggaacg   180
aaagagggtg acttatcaga aaaaccagga gacaccgttc tcactcaagc gaaacctgag   240
ggcgttactg gaaatacgaa ttcacttccg acacctacag aaagaactga agtgagcgag   300
gaaacaagcc cttctagtct ggatacactt tttgaaaaag atgaagaagc tcaaaaaaat   360
ccagagctaa cagatgtctt aaaagaaact gtagatacag ctgatgtgga tgggacacaa   420
gcaagtccag cagaaactac tcctgaacaa gtaaaaggtg gagtgaaaga aaatacaaaa   480
gacagcatcg atgttcctgc tgcttatctt gaaaaagctg aagggaaagg tcctttcact   540
gccggtgtaa accaagtaat tccttatgaa ctattcgctg tgatggtat gttaactcgt   600
ctattactaa aagcttcgga taatgctcct tggtctgaca atggtactgc taaaaatcct   660
gctttacctc ctcttgaagg attaacaaaa gggaaatact ctatgaagt agacttaaat   720
ggcaatactg ttggtaaaca aggtcaagct ttaattgatc aacttcgcgc taatggtact   780
caaacttata aagctactgt taaagtttac ggaaataaag acggtaaagc tgacttgact   840
aatctagttg ctactaagaa tgtagacatc aacatcaatg gattagttgc taaagaaaca   900
gttcaaaaag ccgttgcaga caacgttaaa gacagtatcg atgttccagc agcctaccta   960
gaaaaagcca agggtgaagg tccattcaca gcaggtgtca accatgtgat tccatacgaa  1020
ctcttcgcag gtgatggcat gttgactcgt ctcttgctca aggcatctga caaggcacca  1080
tggtcagata cggcgacgc taaaaaccca gccctatctc cactaggcga aaacgtgaag  1140
accaaaggtc aatacttcta tcaagtagcc ttggacggaa atgtagctgg caaagaaaaa  1200
caagcgctca ttgaccagtt ccgagcaaat ggtactcaaa cttacagcgc tacagtcaat  1260
gtctatggta caaagacgg taaaccagac ttggacaaca tcgtagcaac taaaaaagtc  1320
actattaaca taaacggttt aatttctaaa gaaacagttc aaaaagccgt tgcagacaac  1380
gttaaagaca gtatcgatgt tccagcagca tacctagaaa aagccaaggg tgaaggtcca  1440
ttcacagcag gtgtcaacca tgtgattcca tacgaactct cgcaggtga tggtatgttg  1500
actcgtctct tgctcaaggc atctgacaag gcaccatggt cagataacgg tgacgctaaa  1560
aacccagccc tatctccact aggtgaaaac gtgaagacca aaggtcaata cttctatcaa  1620
ttagccttgg acggaaatgt agctggcaaa gaaaaacaag cgctcattga ccagttccga  1680
gcaaacggta ctcaaactta cagcgctaca gtcaatgtct atggtaacaa agacggtaaa  1740
ccagacttgg acaacatcgt agcaactaaa aaagtcacta ttaacataaa cggtttaatt  1800
tctaaagaaa cagttcaaaa agccgttgca gacaacgtta aggacagtat cgatgttcca  1860
gcagcctacc tagaaaaggc caagggtgaa ggtccattca cagcaggtgt caaccatgtg  1920
attccatacg aactcttcgc aggtgatggc atgttgactc gtctcttgct caaggcatct  1980
gacaaggcac catggtcaga taacggcgac gctaaaaacc cagctctatc tccactaggt  2040
gaaacgtga agaccaaagg tcaatacttc tatcaagtag ccttggacgg aaatgtagct  2100
ggcaaagaaa acaagcgct cattgaccag ttccgagcaa acggtactca aacttacagc  2160
gctacagtca atgtctatgg taacaaagac ggtaaaccag acttggacaa catcgtagca  2220
actaaaaaag tcactattaa gataaatgtt aaagaaacat cagacacagc aaatggttca  2280
```

```
ttatcacctt ctaactctgg ttctggcgtg actccgatga atcacaatca tgctacaggt   2340
actacagata gcatgcctgc tgacaccatg acaagttcta ccaacacgat ggcaggtgaa   2400
aacatggctg cttctgctaa caagatgtct gatacgatga tgtcagagga taaagctatg   2460
ctaccaaata ctggtgagac tcaaacatca atggcaagta ttggtttcct tgggcttgcg   2520
cttgcaggtt tactcggtgg tctaggtttg aaaaacaaaa agaagaaaaa c            2571


SeqID 6
atgaaatcaa taactaaaaa gattaaagca actcttgcag gagtagctgc cttgtttgca     60
gtatttgctc catcatttgt atctgctcaa gaatcatcaa cttacactgt taaagaaggt    120
gatacacttt cagaaatcgc tgaaactcac aacacaacag ttgaaaaatt ggcagaaaac    180
aaccacattg ataacattca tttgatttat gttgatcaag agttggttat cgatggccct    240
gtagcgcctg ttgcaacacc agcgccagct acttatgcgg caccagccgc tcaagatgaa    300
actgtttcag ctccagtagc agaaactcca gtagtaagtg aaacagttgt ttcaactgta    360
agcggatctg aagcagaagc caaagaatgg atcgctcaaa aagaatcagg tggtagctat    420
acagctacaa atggacgtta tatcggacgt taccaattaa cagattcata cctgaacggt    480
gactactcag ctgaaaacca agaacgtgta gcagatgcct acgttgcagg acgttacggt    540
tcatggactg ctgctaaaaa cttctggctt aacaatggct ggtat                    585


SeqID 7
atgaataaga aaaaaatgat tttaacaagt ctagccagcg tcgctatctt aggggctggt     60
tttgttacgt ctcagcctac tttttgtaaga gcagaagaat ctccacaagt tgtcgaaaaa    120
tcttcattag agaagaaata tgaggaagca aaagcaaaag ctgatactgc caagaaagat    180
tacgaaacgg ctaaaaagaa agcagaagac gctcagaaaa agtatgaaga tgatcagaag    240
agaactgagg agaaagctcg aaaagaagca gaagcatctc aaaaattgaa tgatgtggcg    300
cttgttgttc aaaatgcata taaagagtac cgagaagttc aaaatcaacg tagtaaatat    360
aaatctgacg ctgaatatca gaaaaaatta acagaggtcg actctaaaat agagaaggct    420
aggaaagagc aacaggactt gcaaaataaa tttaatgaag taagagcagt tgtagttcct    480
gaaccaaatg cgttggctga gactaagaaa aaagcagaag aagctaaagc agaagaaaaa    540
gtagctaaga gaaaatatga ttatgcaact ctaaaggtag cactagcgaa gaaagaagta    600
gaggctaagg aacttgaaat tgaaaaactt caatatgaaa tttctacttt ggaacaagaa    660
gttgctactg ctcaacatca agtagataat ttgaaaaaac ttcttgctgg tgcggatcct    720
gatgatggca cagaagttat agaagctaaa ttaaaaaaag gagaagctga gctaaacgct    780
aaacaagctg agttagcaaa aaaacaaaca gaacttgaaa aacttcttga cagccttgat    840
cctgaaggta agactcagga tgaattagat aaagaagcag aagaagctga gttggataaa    900
aaagctgatg aacttcaaaa taaagttgct gatttagaaa aagaaattag taaccttgaa    960
atattacttg gaggggctga tcctgaagat gatactgctg ctcttcaaaa taaattagct   1020
gctaaaaaag ctgagttagc aaaaaaacaa acagaacttg aaaaacttct tgacagcctt   1080
gatcctgaag gtaagactca ggatgaatta gataagaag cagaagaagc tgagttggat   1140
aaaaaagctg atgaacttca aaataaagtt gctgatttag aaaaagaaat tagtaacctt   1200
gaaatattac ttggagggc tgattctgaa gatgatactg ctgctcttca aaataaatta   1260
gctactaaaa aagctgaatt ggaaaaaact caaaagaat tagatgcagc tcttaatgag   1320
ttaggccctg atggagatga agaagaaact ccagcgccgg ctcctcaacc agagcaacca   1380
gctcctgcac caaaaccaga gcaaccagct ccagctccaa aaccagagca accagctcct   1440
gcaccaaaac cagagcaacc agctccagct ccaaaaccag agcaaccagc tccagctcca   1500
aaaccagagc aaccagctaa gccggagaaa ccagctgaag agcctactca accagaaaaa   1560
ccagccactc caaaaacagg ctggaaacaa gaaaacggta tgtggtattt ctacaatact   1620
gatggttcaa tggcaatagg ttggctccaa aacaacggtt catggtacta cctaaacgct   1680
aacggcgcta tggcaacagg ttgggtgaaa gatggagata cctggtacta tcttgaagca   1740
tcaggtgcta tgaaagcaag ccaatggttc aaagtatcag ataaatggta ctatgtcaac   1800
agcaatggcg ctatggcgac aggctggctc caatacaatg gctcatggta ctacctcaac   1860
gctaatggtg atatggcgac aggatggctc caatacaacg gttcatggta ttacctcaac   1920
gctaatggtg atatggcgac aggatgggct aaagtcaacg gttcatggta ctacctaaac   1980
gctaacggta tatggcgac aggtgacgc aaagtcaacg gttcatggta ctacctaaac   2040
gctaacggtt caatggcaac aggttgggtg aaagatggag atacctggta ctatctttgaa   2100
gcatcaggtg ctatgaaagc aagccaatgg ttcaaagtat cagataaatg gtactatgtc   2160
aatggcttag gtgcccttgc agtcaacaca actgtagatg gctataaagt caatgccaat   2220
ggtgaatggg tt                                                        2232


SeqID 8
atgaaaaaaa tagttcttgt tagtctagct ttccttttg tcctggttgg ttgcggacag     60
aaaaaagaaa ctggaccagc tacaaaaaca gaaaaagata cgcttcagtc ggcattgcca    120
gttattgaaa atgctgagaa gaatacagtt gtaactaaga cttggtctt gcccaagtca    180
gatgatggta gccagcagac acaaacaatt acttacaaag acaagacttt tttgagtcta    240
gctatccaac aaaaacgtcc agtctctgat gagttgaaga cttatattga ccaacatgga    300
gtggaggaaa ctcaaaaagc tcttcttgaa gcggaggaga aggataagtc tatcattgaa    360
gctcgtaaat tggcaggttt caaacttgaa acaaaactat tgagcgcaac ggaacttcaa    420
acaacgacta gttttgattt tcaagttctg gatgtcaaga aggcttccca gttggaacat    480
ctgaagaata ttggtttgga aaatctttttg aaaaatgaac caagcaaata tatttcagat    540
```

```
agattggcaa atggcgcgac agaacaa                                    567

SeqID 9
atgtttgaag tagaagaatg gctccatagt cggattggtt tgaattttcg atcaggtttg      60
ggtcgaatgc agcaagcggt ggatttgtta ggaaatcctg agcagtctta ccctattatc     120
cacgtaacag ggactaatgg gaaaggatct accattgctt ttatgaggga attatttatg     180
gggcatggca aaaagttgc gacctttacc tcccctcata tcgtctctat caatgaccga     240
atctgcatta atgggcagcc tatagcagac gcagactta tccgtttgac tgatcaggtc     300
aaggagatgg agaaaacgct tctgcaaact cctgcccagt tgtccttttt tgaattgctg     360
accttggttg cttttctta ttttagggag caggaggtgg atttggtttt attagaagtg      420
ggaattggtg gcttacttga cacgaccaat gtggtaactg gagagtttgc tgtcatcacc     480
tccattgggc ttgaccatca agaaaccttg ggtgatagtc tagaagcaat tgcagagcag     540
aaagctggta tttttcaaggc tggtaagaag gcagtgattg cgaaattgcc tccagaagct     600
aggcttgcct gtcagaaaaa agccgaatct ttagctgtta accttttatca ggcaggtcaa     660
gatttttttaa tgctgaatgg tgattttttca agctctttac taaatctttc acagctgaac     720
ataggcttag aaggagtctca tcagcaggag aatgcagcct ggcgttgca aacttttctt     780
cttttatga gagaaagaaa ggaagctgtt gatgaacagg ctgtaagaaa ggccttggaa     840
cagacccatt gggctggtcg cttggagcgt attcgcccac agatttattt ggatggtgct     900
cataacctcc ctgccttgac tcgcttggct gagtttatca aagaaaaaga gcaggaaggc     960
tatcgacctc aaatcctctt tggatccttg aaacgtaagg attatcaagg gatgttgggt    1020
tatctgactg aaaaattgcc tcaggtggaa ctcaaggtga ccggctttga ctatcagggg    1080
gctttggacg aaagggatgt gacaggttac gatatagttt cttcttaccg agaatttatc    1140
agcgattttg aagaaagggc agacgctcaa gacttgctgt tcgttacagg gtctctctat    1200
tttatctcag aagtacgggg ctacctgctg gaccgtgagc agataaat                 1248

SeqID 10
gtgggaattc gtgtttataa accaacaaca aacggtcgcc gtaatatgac ttctttggat      60
ttcgctgaaa tcacaacaag cactcctgaa aaatcattgc ttgttgcatt gaagagcaag     120
gctggtcgta acaacaacgg tcgtatcaca gttcgtcacc aaggtggtgg acacaaacgt     180
ttctaccgtt tggttgactt caaacgtaat aaagacaacg ttgaagcagt tgttaaaaca     240
atcgagtacg atccaaaccg ttctgcaaac atcgctcttg tacactacac tgacggtgtg     300
aaagcataca tcatcgctcc aaaaggtctt gaagtaggtc aacgtatcgt ttcaggtcca     360
gaagcagata tcaaagtcgg aaacgctctt ccacttgcta acatcccagt tggtactttg     420
attcacaaca tcgagttgaa accaggtcgt ggtggtgaat tggtacgtgc tgctggtgca     480
tctgctcaag tattgggttc tgaaggtaaa tatgttcttg ttcgtcttca atcaggtgaa     540
gttcgtatga ttcttggaac ttgccgtgct acagttggtg ttgtcggaaa cgaacaacat     600
ggacttgtaa accttggtaa agcaggacgt agccgttgga aaggtatccg cccaacagtt     660
cgtggttctg taatgaaccc taacgatcac ccacacggtg gtggtgaagg taaagcacca     720
gttggtcgta aagcaccatc tactccatgg ggcaaacctg ctcttggtct taaaaactcgt     780
aacaagaaag cgaaatctga caaacttatc gttcgtcgtc gcaacgagaa a              831

SeqID 11
atggctaaaa aatcaatggt agctagagag gctaaacgcc aaaaaattgt tgaccgttat      60
gctgaaaaac gtgctgcatt aaaggcggca ggggactacg aaggtttatc taaattacct     120
cgcaacgcct caccgactcg tttacataat cgttgtaggg ttacggggcg cccacattca     180
gtttaccgca aatttggtct gagtcgtatc gcttttcgcg aacttgcgca taaaggtcaa     240
attcctggtg taacaaaagc atcttgg                                         267

SeqID 12
atggatatta gacaagttac tgaaaccatc gccatgattg aggagcaaaa cttcgatatt      60
agaaccatta ccatggggat ttctcttttg gactgtatcg atccagatat caatcgtgct     120
gcggagaaaa tctatcaaaa aattacgaca aaggcggcta atttagtagc tgttggtgat     180
gaaattgcgg ctgagttggg aattcctatc gttaataagc gtgtatcggt gacacctatt     240
tctctgattg gggcagcgac agatgcgacg gactacgtgg ttctggcaaa agcgcttgat     300
aaggctgcga aagagattgg tgtggacttt attggtggtt tttctgcctt agtacaaaaa     360
ggttatcaaa agggagatga gattctcatc aattccattc ctcgcgcttt ggctgagacg     420
gataaggtct gctcgtcagt caatatcggc tcaaccaagt ctggtattaa tatgacggct     480
gtggcagata tgggacgaat tatcaaggaa acagcaaatc tttcagatat gggagtggcc     540
aagttggttg tattcgctaa tgctgttgag gacaatccat ttatggcggg tgcctttcat     600
ggtgttgggg aagcagatgt tatcatcaat gtcggagttt ctggtcctgg tgttgtgaaa     660
cgtgctttgg aaaaagttcg tggacagagc tttgatgtag tagccgaaac agttaagaaa     720
actgccttta aaatcactcg tatcggtcaa ttggttggtc aaatggccag tgagagactg     780
ggtgtggagt ttggtattgt ggacttgagt ttggcaccaa cccctgcggt tggagactct     840
gtggcacgtg tccttgagga atggggcta gaaacagttg gcacgcatgg aacgacggct     900
gccttggccc tcttgaacga ccaagttaaa aagggtggag tgatggcctg caaccaagtc     960
ggtggtttat ctggtgcctt tatccctgtt ctgaggatg aaggaatgat tgctgcagtg    1020
caaaatggct ctcttaattt agaaaaacta gaagctatga cggctatctg ttctgttgga    1080
ttggatatga ttgccatccc agaagatacg cctgctgaaa ctattgcggc tatgattgcg    1140
```

```
gatgaagcag caatcggtgt tatcaacatg aaaacaacag ctgttcgtat cattcccaaa      1200
ggaaaagaag gcgatatgat tgagtttggt ggtctattag gaactcgcacc cgttatgaag     1260
gttaatgggg cttcgtctgt cgacttcatc tctcgcggtg gacaaatccc agcaccaatt      1320
catagtttta aaaat                                                       1335

SeqID 13
atggtaaata cagaagtagc aagaacaaca atcaagacag aatattttgg cagccttact      60
gaaaggatga acaaatatcg agaagatgtt ttaaataaaa aaccttatat tgatgctgag      120
agagcagttc tagcaacacg cgcctatgaa cgatacaagg aacaacctaa tgtcctaaaa      180
cgtgcatata tgctgaaaga aattttggaa aatatgacta tctatattga agaagaatct      240
atgattgcgg gaaatcaagc ttcttccaat aaagatgctc ctattttttcc ggaatatacg      300
ctagaatttg ttctcaatga gttggatctt tttgaaaagc gtgatggaga tgttttctat      360
attacagaag aaacaaaaga acaacttaga agtattgctc cgttttggga aaataataat      420
ttacgtgcta gagctggtgc cttattacct gaagaagtgt ctgtttatat ggaaacagga      480
ttcttcggta tggaaggtaa gatgaattct ggagatgctc acttagcagt taactatcag      540
aaactttttgc aatttggttt aagaggtttt gaagagcggg ctcgtaaagc aaaagtagct      600
ctagatttaa cagatccagc aagtattgat aaatatcatt tttacgactc tatatttatc      660
gtaatcgatg ctattaaagt atatgcaaag cgctttgttg ctcttgctaa aagtttagcc      720
gaaaatgcaa atcctaaacg taagaaagaa ttacttgaga ttgcagatat ttgctctaga      780
gtcccatatg aaccggcaac tactttttgca gaagctattc aatcagtttg gtttattcaa      840
tgtatttttac aaattgaatc taatggccac tctctttcat atggccgttt tgatcaatat      900
atgtatccat atatgaaggc tgatttagaa agtggtaaag aaacagaaga tagcattgtt      960
gaacgtctga caaatctttg gattaagaca attacaatta ataaggttcg cagtcaatca     1020
catacatttt cttcagcagg aagtcctttta tatcaaaatg ttacaattgg tggacagact     1080
cgagataaga aggatgctgt taacccatta tcttatttgg tattaaaatc agttgcacaa     1140
acccatctac cgcaacctaa tctaactgta cgttaccatg caggtttaga tgctcgtttc     1200
atgaatgagt gtattgaagt gatgaaactt ggtttttggta tgcctgcatt taataatgat     1260
gagattatta ttccttcttt tattgcaaga ggattattgg aagatgatgc ttatgattac     1320
agtgccattg gatgtgttga aacggcagtt ccaggggaaat ggggctatcg ttgcacaggt     1380
atgagttata tgaacttccc taaggttcta cttatcacga tgaatgatgg aattgatccg     1440
gcttcgggta aacggtttgc accaagcttt ggtcgtttta aggatatgaa gaactttct      1500
gaattagaaa atgcttggga taaaaacacta agatatttga cacgaatgag tgttattgtt     1560
gaaaattcta ttgatttatc attggaacga gaagttcctg atattctatg ttcagcattg     1620
actgatgatt gtattggtcg tggaaaacac cttaaagaag gtggagcagt atatgattat     1680
atatcaggat tgcaagttgg aattgcaaat ttgtcggatt cattagctgc aattaaaaaa     1740
ttggtgtttg aggaagaacg tataagccca agtcagcttt ggcatgcact ggaaacagat     1800
tatgccggag aagaaggtaa ggtcattcaa gaaatggttga ttcatgatgc acctaagtat     1860
ggtaatgatg atgattatgc tgacaaattg gttactgctg cttatgacat ttatgttgat     1920
gaaattgcta aatatcctaa tacacgttat ggaagagggc ctattggagg aattcgttat     1980
tcaggaacat cttctatctc agccaacgta gggcagggac gtggaacatt agcaactcca     2040
gatggacgca acgcgggtac accgttagca gagggttgtt caccatcaca taatatggat     2100
caacacggcc ctacatctgt tttaaaatct gtttcaaaat taccaacaga tgaaatcgta      2160
ggtggggttc tcttaaatca gaaagtaaat cctcaaacgt tagccaaaga agaagataaa     2220
ttaaaactaa ttgctttgtt acgaacattc tttaatcgtt tacatgggta ccatattcaa     2280
tacaatgttg tttccagaga gacgctgatt gacgctcaga aacatcctga aaaacacaga     2340
gacttaattg ttcgtgttgc aggatactct gcattcttca atgttcttttc taaggcaacc     2400
caagatgaca ttataggacg tactgagcat actttg                                2436

SeqID 14
atgtcacaag cacaatatgc aggtactgga cgtcgtaaaa acgctgttgc acgcgttcgc      60
cttgttccag gaactggtaa aatcactgtt aacaaaaaag atgttgaaga gtacatccca      120
cacgctgacc ttcgtcttgt catcaaccaa ccattcgcag ttacttcaac tgtaggttca      180
tacgacgttt tcgttaacgt tataggtggt ggatacgctg gtcaatcagg agctatccgt      240
cacggtatcg ctcgtgccct tcttcaagta gaccagagact tccgcgattc attgaaacgc      300
gcaggacttc ttacacgtga ctcacgtaaa gttgaacgta agaaaccagg tcttaagaaa      360
gctcgtaaag catcacaatt tagtaaacgt                                       390

SeqID 15
ttggagaaga aactgaccat aaaagacatt gcggaaatgg ctcagacctc gaaaacaacc      60
gtgtcatttt acctaaacgg gaaatatgaa aaaatgtccc aagagacacg tgaaaagatt      120
gaaaaagtta ttcatgaaac aaattacaaa ccgagcattg ttgcgcgtag cttaaactcc      180
aaacgaacaa aattaatcgg tgttttgatt ggtgatatta ccaacagttt ctcaaaccaa      240
attgttaagg gaattgagga tatcgccagc cagaatggct accaggtaat gataggaaat      300
agtaattaca gccaagagag tgaggacggg tatattgaaa gcatgcttct cttgggagta      360
gacggcttta ttattcagcc gacctctaat ttccgaaaat attctcgtat catcgatgag      420
aaaaagaaga aaatggtctt ttttgatagt cagctctatg aacaccggac tagctgggtt      480
aaaaccaata actatgatgc cgtttatgac atgacccagt cctgtatcga aaaaggttat      540
gaacattttc tcttgattac agcggatacg agtcgtttga gtactcggat tgagcgggca      600
```

```
agtggttttg tggatgcttt aacagatgct aatatgcgtc acgccagtct aaccattgaa      660
gataagcata cgaatttgga acaaattaag gaatttttac aaaaagaaat cgatcccgat      720
gaaaaaactc tggtatttat ccctaactgt tgggccctac ctctagtctt taccgttatc      780
aaagagttga attataactt gccacaagtt gggttgattg ttttgacaa tacggagtgg       840
acttgctttt cttctccaag tgtttcgacg ctggttcagc cctcctttga ggaaggacaa      900
caggctacaa agattttgat tgaccagatt gaaggtcgca atcaagaaga aaggcaacaa      960
gtcttggatt gtagtgtgaa ttggaaagag tcgactttc                            999
```

SeqID 16
```
atgaataaag gattatttga aaaacgttgt aaatatagta ttcggaaatt ttcattaggt       60
gttgcttctg ttatgattgg agctgcattc tttgggacaa gtccggttc tgcagatagc       120
gtgcagtctg gttccacggc gaacttacca gctgattag ctactgctct tgcaacagca       180
aaagagaatg atgggcgtga ttttgaagcg cctaaggtgg gagaagacca aggttctcca       240
gaagttacag atggacctaa gacagaagaa gaactattag cacttgaaaa agaaaaaccg       300
gctgaagaaa aaccaaaaga ggataaacct gcagctgcta aacctgaaac acctaagacg       360
gtaaccctga aatggcaaac ggtagcgaat aaagagcaac agggaacagt cactatccga       420
gaagaaaaag gtgtccgcta caaccaacta tcctcaactg ctcaaaatga taacgcaggc       480
aaaccagccc tgtttgaaaa gaagggcttg accgttgatg ccaatggaaa tgcaactgtt       540
gatttaacct tcaaagatga ttctgaaaag ggcaaatcac gctttggtgt cttttttgaaa       600
tttaaagata ccaagaataa tgttttttgtc ggttatgaca aggatggctg gttctgggag       660
tataaatctc caacaactag cacttggtat agaggtagtc gtgttgctgc tcctgaaaca       720
ggatcaacaa accgtctctc tatcactctc aagtcagacg gtcagctaaa tgccagcaat       780
aatgatgtca atctctttga cacagtgact ctaccagctg cggtcaatga ccatcttaaa       840
aatgagaaga agattcttct caaggcgggc tcttatgacg atgagcgaac agttgttagc       900
gttaaaacgg ataaccaaga gggggtaaaa acagaggata cccctgctgg aaaagaaaca       960
ggtcctgaag ttgatgatag caaggtgact tatgacacga ttcagtctaa ggtcctcaaa      1020
gcagtgattg accaagcctt ccctcgtgtc aaggaataca gcttgaacgg gcatactttg      1080
ccaggacagg tgcaacagtt caaccaagtc tttatcaata accaccgaat caccccctgaa      1140
gtcacttata agaaaatcaa tgagacaaca gcagagtact tgatgaagct cgcgatgat       1200
gctcacttaa tcaatgcgga aatgacagta cgcttgcaag ttgtagacaa tcaattgcac      1260
tttgatgtga ctaagattgt caaccacaat caagtcactc caggtcaaaa gattgatgac      1320
gaaagcaaac tactttcttc tattagtttc ctcggcaatg ctttagtctc tgtttctagt      1380
aatcaaactg gtgctaagtt tgatggggca accatgtcaa acaatacgac tgtcagcgga      1440
gatgatcata tcgatgtaac caatccaatg aaggatttgg ctaagggtta catgtatgga      1500
tttgtttcta cagataagct tgctgctggt gtttggagta actctcaaaa cagctatggt      1560
ggtggttcga atgactggac tcgtttgaca gcttataaag aaacagtcgg aaatgccaac      1620
tatgtaggaa tccacagctc tgaatggcaa tgggaaaaag cttataaggg cattgttttc      1680
ccagaataca cgaaggaact tccaagtgct aaggttgtta tcactgaaga tgccaatgca      1740
gacaagaacg ttgattggca agatggtgcc attgcttatc gtagcattat gaacaatcct      1800
caaggttggg aaaaagttaa ggatatcaca gcttaccgta tcgcgatgaa ctttggttct      1860
caagcacaaa acccattcct tatgaccttg gatggtatca agaaaatcaa tctccataca      1920
gatggtcttg ggcaaggtgt tctccttaaa ggatatggta gcgaaggcca tgactctggt      1980
cacttgaact atgctgatat tggtaagcgt atcggtggtg tcgaagact caagacccta      2040
attgagaagg ctaagaaata tggagctcat ctaggtatcc acgttaacgc ttcagaaact      2100
tatcctgagt ctaaatactt caatgaaaaa attctccgta agaatccaga tggaagctat      2160
agctatggtt ggaactggct agatcaaggt atcaacattg atgctgccta tgacctagct      2220
catggtcgtt tggcacgttg ggaagatttg aagaaaaaac ttggtgacgg tctcgacttt      2280
atctatgtgg acgtttgggg taatggtcaa tcaggtgata acggtgcctg ggctacccac      2340
gttcttgcta aagaaattaa caaacaaggc tggcgctttg cgatcgagtg gggccatggt      2400
ggtgagtacg actctacctt ccatcactgg gcagctgact tgacctacgg tggctacacc      2460
aataaaggta tcaacagtgc catcacccgc tttatccgta accaccaaaa agatgcttgg      2520
gtaggggact acagaagtta tggttggtgca gccaactatc cactgctagg tggctacacgc      2580
atgaaagact ttgaaggctg gcaggaaga agtgactaca atggctatgt aaccaactta      2640
tttgcccatg acgtcatgac taagtacttc caacacttca ctgtaagtaa atgggaaaat      2700
ggtacaccgg tgactatgac cgataacggt agcacctata aatggactcc agaaatgcga      2760
gtggaattgg tagatgctga caataataaa gtagttgtaa ctcgtaagtc aaatgatgtc      2820
aatagtccac aatatcgcga acgtacagta acgctcaacg gacgtgtcat ccaagatggt      2880
tcagcttact tgactccttg gaactgggat gcaaatggta gaaactttc tactgataag      2940
gaaaagatgt actacttcaa tacgcaggcc ggtgcaacaa cttggaccct tccaagcgat      3000
tgggcaaaga gcaaggttta cctttacaag ctaactgacc aaggtaagac agaagagcaa      3060
gaactaactg taaaagatgg taaaattacc ctagatcttc tagcaaatca accatacgtt      3120
ctctatcgtt cgaaacaaac taatcctgaa atgtcatgga gtgaaggcat gcacatctat      3180
gaccaaggat ttaatagcgg taccttgaaa cattggacca tttcaggcga tgcttctaag      3240
gcagaaattg tcaagtctca aggggcaaac gatatgcttc gtattcaagg aaacaaagaa      3300
aaagttagtc tcactcagaa attaactggc ttgaaaccaa ataccaagta tgccgtttat      3360
gttggtgtag ataaccgtag taatgccaag gcaagtatca ctgtgaatac tggtgaaaaa      3420
gaagtgacta cttataccaa taagtctctc gcgctcaact atgttaaggc ctacgcccac      3480
aatacacgtc gtgacaatgc tacagttgac gatacaagtt acttccaaaa catgtacgcc      3540
```

```
ttctttacaa ctggagcgga cgtctcaaat gttactctga cattgagtcg tgaagctggt   3600
gatcaagcaa cttactttga tgaaattcgt acctttgaaa acaattcaag catgtacgga   3660
gacaagcatg atacaggtaa aggcaccttc aagcaagact ttgaaaatgt tgctcagggt   3720
atcttcccat ttgtagtggg tggtgtcgaa ggtgttgaag ataaccgcac tcacttgtct   3780
gaaaaacaca atccatatac acaacgtggt tggaatggta agaaagtcga tgatgttatc   3840
gaaggaaatt ggtcactcaa gacaaatgga ctagtgagcc gtcgtaactt ggtttaccaa   3900
accatcccac aaaacttccg ttttgaagca ggtaagacct accgtgtaac ctttgaatac   3960
gaagcaggat cagacaatac ctatgctttt ctagtcggta agggagaatt ccagtcaggt   4020
cgtcgtggta ctcaagcaag caacttggaa atgcatgaat tgccaaatac ttggacagat   4080
tctaagaaag ccaagaaggc aaccttcctt gtgacaggtg cagaaacagg cgatacttgg   4140
gtaggtatct actcaactgg aaatgcaagt aatactcgtg gtgattctgg tggaaatgcc   4200
aacttccgtg gttataacga cttcatgatg gataatcttc aaatcgaaga aattacccta   4260
acaggtaaga tgttgacaga aaatgctctg aagaactact tgccaacggt tgccatgact   4320
aactacacca aagagtctat ggatgctttg aaagaggcgg tctttaacct cagtcaggcc   4380
gatgatgata tcagtgtgga agaagcgcgt gcagagattg ccaagattga agctttgaag   4440
aatgctttgg ttcagaagaa gacggctttg gtagcagatg actttgcaag tcttacagct   4500
cctgtcaggc ctcaagaagg tcttgcaaat gcctttgagg gcaatgtgtc tagtctatgg   4560
catacatctt ggaatggtgg agatgtaggc aagcctgcaa ctatggtctt gaaagaacca   4620
actgaaatca caggacttcg ctatgttccg cgtggatcag gttcaaatgg taacttgcga   4680
gatgtgaaac ttgttgtgac agatgagtct ggcaaggagc ataccttttac tgcaactgat   4740
tggccaaata acaacaaacc aaaagatatt gacttggta agacaatcaa ggctaagaaa   4800
attgtcctta ctggtaccaa gacatacgga gatggtggag ataaataccca atctgcagcg   4860
gaacttatct ttactcgtcc acaggtagca gaaacacctc ttgacttgtc aggctatgaa   4920
gcagctttgg ttaaggctca gaaattaaca gacaaagaca atcaagagga agtagctagc   4980
gttcaggcaa gcatgaaata tgcgacggat aaccatctct tgacggaaag aatggtggaa   5040
tactttgcag attatctcaa ccaattaaaa gattctgcta cgaaaccaga tgctccaact   5100
gtagagaaac ctgagtttaa acttagatct ttagcttccg agcaaggtaa gacgccagat   5160
tataagcaag aaatagctag accagaaaca cctgaacaaa tcttgccagc aacaggtgag   5220
agtcaatctg acacagccct catcctagca agtgttagtc tagccctatc tgctctcttt   5280
gtagtaaaaa cgaagaaaga c                                              5301
```

SeqID 17
```
atgaacaaac caacgattct gcgcctaatc aagtatctga gcattagctt cttaagcttg     60
gttatcgcag ccattgtctt aggcggagga gttttttttct actacgttag caaggctcct    120
agcctatccg agagtaaact agttgcaaca acttctagta aatctacga caataaaaat    180
caactcattg ctgacttggg ttctgaacgc cgcgtcaatg cccaagctaa tgatattccc    240
acagatttgg ttaaggcaat cgtttctatc gaagaccatc gcttcttcga ccacaggggg    300
attgatacca tccgtatcct gggagctttc ttgcgcaatc tgcaaagcaa ttccctccaa    360
ggtggatcaa ctctcaccca acagttgatt aagttgactt acttttcaac ttcgacttcc    420
gaccagacta tttctcgtaa ggctcaggaa gcttggttag cgattcagtt agaacaaaaa    480
gcaaccaagc aagaaatctt gacctactat ataaataagg tctacatgtc taatgggaac    540
tatggaatgc agacagcagc tcaaaactac tatggtaaag acctcaataa tttaagttta    600
cctcagttag ccttgctggc tggaatgcct caggcaccaa accaatatga ccccctattca    660
catccagaag cagcccaaga ccgccgaaatc ttggtcttat ctgaaatgaa aaatcaaggc    720
tacatctctg ctgaacagta tgagaaagca gtcaatacac caattactga tggactacaa    780
agtctcaaat cagcaagtaa ttaccctgct tacatggata attacctcaa ggaagtcatc    840
aatcaagttg aagaagaaac aggctataac ctactcacaa ctgggatgga tgtctacaca    900
aatgtagacc aagaagctca aaaacatctg tgggatattt acaatacaga cgaatacgtt    960
gcctatccag acgatgaatt gcaagtcgct tctaccattg ttgatgtttc taacggtaaa   1020
gtcattgccc agctaggagc acgccatcag tcaagtaatg tttccttcgg aattaaccaa   1080
gcagtagaaa caaaccgcga ctgggatca actatgaaac cgatcacaga ctatgctcct   1140
gccttggagt acggtgtcta cgattcaact gctacatgta ttcaacttga gccctataac   1200
taccctggga caaatactcc tgtttataac tgggatagg gctacttgg caacatcacc   1260
ttgcaatacg ccctgcaaca atcgcgaaac gtcccagccg tggaaactct aaacaaggtc   1320
ggactcaacc gcgccaagac tttcctaaat ggtctaggaa tcgactaccc aagtattcac   1380
tactcaaatg ccatttcaag taacacaacc gaatcagaca aaaaatatgg agcaagtagt   1440
gaaagatgg ctgctgctta cgctgccttt gcaaatggtg gaacttacta taaaccaatg   1500
tatatccata aagtcgtctt tagtgatggg agtgaaaaag agttctctaa tgtcggaact   1560
cgtgccatga aggaaacgac agcctatatg atgaccgaca tgatgaaaac agtcttgact   1620
tatggaactg gacgaaatgc ctatcttgct tggctccctc aggctggtaa aacaggaacc   1680
tctaactata cagacgagga aattgaaaac cacatcaaga cctctcaatt tgtagcacct   1740
gatgaactat ttgctggcta tacgcgtaaa tattcaatgt ctgtatggac aggctattct   1800
aaccgtctga caccacttgt aggcaatggc cttacggtcg ctgccaaagt ttaccgctct   1860
atgatgacct acctgtctga aggaagcaat ccagaagatt ggaatatacc agaggggctc   1920
tacagaaatg gagaattcgt attttaaaaat ggtgctcgtt tacgtggaa ctcacctgct   1980
ccacaacaac ccccatcaac tgaaagttca agctcatcat cagatagttc aacttcacag   2040
tctagctcaa ccactccaag cacaaataat agtacgacta ccaatcctaa caataatacg   2100
caacaatcaa atacaacccc tgatcaacaa aatcagaatc ctcaaccagc acaacca      2157
```

```
SeqID 18
atgagtaaaa aaagacgaaa tcgtcataaa aaagaaggtc aagaaccgca atttgatttt      60
gatgaagcaa aagagctaac agttggtcaa gctattcgta aaaatgaaga agtggaatca     120
ggagtcttgc ctgaggattc cattttggac aagtatgtta agcaacacag agatgaaatt     180
gaggcggata agtttgcgac tcgtcaatac aaaaaagagg agttcgttga aactcagagt     240
ctggatgatt taattcaaga gatgcgtgag gctgtagaga agtcagaagc ttcttcggag     300
gaagttccat cttctgaaga catcttacta cccttgcctc tggacgatga ggagcaaggc     360
ttggatcctc tattgctaga tgatgaaaat ccaacagaaa tgactgaaga agtggaagag     420
gagcaaaacc tttctcgtct ggatcaagag gactcagaaa agaaaagtaa aaaaggcttt     480
attttgaccg ttttggcgct tgtatcagta attatttgtg tcagtgctta ttatgtctac     540
cgtcaagtgg ctcgttcgac taaggaaatt gaaacttctc aatcaactac agccaatcaa     600
tcggatgtgg atgatttaa tacactttat gacgcctttt acacagatag caataaaacg     660
gctttgaaaa atagccagtt tgataaactg agtcaactca agactttact tgataagctg     720
gaaggtagtc gtgaacatac gcttgccaaa tctaaatatg atagtctagc aacgcaaatc     780
aaggctattc aagatgtcaa tgctcaattt gagaaaccag ctattgtgga tggtgtgttg     840
gataccaatg ccaaagccaa atcggatgct aaatttacgg atattaaaac tggaaatacg     900
gagcttgata aagtgctaga taaggctatc agtcttggta agagccagca aacaagtact     960
tctagctcaa gttcaagtca aactagcagc tcaagttcaa gtcaagcaag ttcaaatacg    1020
actagtgagc caaaaccaag tagttcaaat gagactagaa gtagtcgcag tgaagtcaat    1080
atgggtctct cgagtgcagg ggttgctgtt caaagaagtg ccagtcgtgt tgcctataat    1140
cagtctgcta ttgatgatag taataactct gcctgggatt ttgcggatgg tgtcttggaa    1200
caaattctag cgacttcacg ttcacgtggc tatatcactg gagaccaata tatccttgaa    1260
cgtgtcaata tcgttaacgg caatggttat tacaacctct acaagccaga tggaacctat    1320
ctctttaccc ttaactgtaa gacaggctac tttgtcggaa atggcgctgg tcatgcggat    1380
gacttagatt ac                                                        1392

SeqID 19
atgaagcttt tgaaaaaaat gatgcaagtc gcattagcag tctttttctt tggtttgcta      60
gctacaaata cggtatttgc gaataccaca ggtggccgat ttgttgataa ggataataga     120
aaatattatg taaaagatga tcataaagca atctattggc ataaaataga cggtaaaact     180
tactattttg gtgatattgg agagatggtt gtcggttggc aatacttaga aattcctgga     240
acaggttatc gtgataattt attcgataac caaccagtta atgaaattgg ccttcaggag     300
aagtggtact attttggaca agatggtgct ttgctagaac aaacagataa acaagtacta     360
gaggcaaaaa cgtctgaaaa tacaggaaaa gtatacggtg aacaatatcc tctatctgct     420
gaaaagagaa cttattattt tgataataat tatgctgtaa agacaggctg gatttatgaa     480
gagggtcatt ggtattattt aaataagcta ggaaattttg gcgatgattc ttacaatcca     540
ctaccaattg gtgaagttgc taaggggttgg actcaagatt ttcatgttac tattgacatt     600
gatagaagca aacctgctcc atggtactac ctagatgctt caggtaagat gcttacagat     660
tggcaaaaag taaacggaaa atggtattat tttggctcct ctggttctat ggcaacaggt     720
tggaaatatg tacgaggcaa atggtattac ttagataata aaaatggtga tatgaaaaca     780
ggatggcaat accttggtaa caagtggtac tacctccgtt catcaggagc tatggtaact     840
ggctggtatc aagatggttc aacttggtac tatttagatc cttctaatgg agatatgaaa     900
ataggttgga caaaagtaaa tggaaaatgg tattatctca attcaaatgg agcaatggtt     960
acaggtagcc aaactatcga tggtaaagtt tataatttcg cctcatctgg tgagtggatt    1020

SeqID 20
atgaaaattt tgaaaaaaac tatgcaagtt ggactgacag tattttttctt tggtttgcta      60
gggaccagta cagtatttgc agatgattct gaaggatggc agtttgtcca agaaaacgga     120
agaacctact acaaaaaggg ggacctcaaa gaaacctact ggcgagtgat tgatggtaag     180
tactattatt ttgattctct atctggagag atggttgtcg gctggcaata tatcccgttt     240
ccatctaaag gtagtacaat tggtccttac ccaaatggta tcagattaga aggttttcca     300
aagtcagagt ggtactactt cgataaaaat ggagtgctac aagagtttgt tggttggaaa     360
acattagaga ttaaaactaa agacagtgtt ggaagaaagt acggggaaaa acgtgaagat     420
tcagaagata aagaagagaa gcgttattat acgaactatt actttaatca aaatcattct     480
ttagagacag gttggcttta tgatcagtct aactggtatt atctagctaa gacggaaatt     540
aatggagaaa actaccttgg tggtgaaaga cgtgcgggt ggataaacga tgattcgact     600
tggtactacc tagatccaac aactggtatt atgcaaacag gttggcaata tctaggtaat     660
aagtggtact acctccgttc ctcaggagca atggccactg ctggtatcac ggaaggtacc     720
acttggtatt atttagacca cccaaatggc gatatgaaaa caggttggca aaaccttggg     780
aacaaatggt actatctccg ttcatcagga gctatggcaa ctggttggta tcaagatggt     840
tcaacttggt actacctaaa tgcaggtaat ggagacatga agacaggttg gttccaggtc     900
aatggcaact ggtactatgc ttatagctca ggtgctttgg cagtgaatac gaccgtagat     960
ggctattctg tcaactataa tggcgaatgg gttcgg                               996

SeqID 21
atggttttat ctaagtatta tggagtagcc gatggaatga atgtagaagg gaggggaagt      60
gcgaatttta ttaaagataa tgtgttaatt acagcggctc acaactacta cagacatgac     120
```

```
tatgggaaag aagcggatga tatttatgtt cttccggctg ttagtccaag tcaagaacca      180
tttggaaaga tcaaagtaaa ggaagttcgt tatttgaagg aatttagaaa tttaaattct .    240
aaggatgcaa gggaatatga cttggcttta ttaattctag aagagcccat tggtgcaaaa      300
ttagggactt tgggtcttcc tactagtcaa aaaaatttga caggaataac tgtgactatc      360
acaggctatc catcatataa ttttaaaatt catcaaatgt atacagataa aaaacaagtt      420
ttaagtgatg atggcatgtt cttggattac caagttgata ctttagaggg gtctagtgga      480
tctacagttt atgatgctag tcaccgtgta gtaggagtgc atactttagg agatggagct      540
aatcaaatta acagtgcagt taaattaaat gaacgaaatt tgccatttat ttattcggtt      600
cttaaaggtt actctcttga aggatggaag aaaataaatg gtagttggta ccattataga      660
caacatgata aacaaacggg ttggcaggag ataaatgata cttggtatta tttagacagt      720
tccggtaaga tgcttacaga ttggcaaaaa gtaaatggaa aatggtatta tctcaattca      780
aatggagcaa tggttacagg tagccaaact atcgatggta aagtttataa cttcgcttca      840
tctggtgagt ggatt                                                       855


SeqID 22
ttgatgaaaa aaactttttt cttactggtg ttaggcttgt tttgccttct tccactctct       60
gttttttgcca ttgatttcaa gataaactct tatcaagggg atttgtatat tcatgcagac      120
aatacggcag agtttagaca gaagatagtt taccagtttg aggaggactt taagggccaa      180
atcgtgggac ttggacgtgc tggtaagatg cctagcgggt ttgacattga ccctcatcca      240
aagattcagg ccgcgaaaaa cggtgcagaa ctagcagatg tgactagcga agtaacagaa      300
gaagcggatg gttatactgt gagagtctat aatccaggtc aggagggcga catagttgaa      360
gttgacctcg tctggaactt aaaaaattta cttttccttt atgatgatat cgctgaatta      420
aattggcaac ctctgacaga tagttcagag tctattgaaa agtttgaatt tcatgtaagg      480
ggagacaagg gggctgaaaa actctttttc catcaggga aacttttag agagggaacg       540
attgaaaaga gtaaccttga ttatactatc cgtttagaca atcttccgtc taagcgtgga      600
gttgagttgc atgcctattg gcctcggacc gattttgcta gcgctaggga tcagggattg      660
aaagggaatc gtttagaaga gtttaataag atagaagact cgattgttag agaaaaagat      720
cagagtaaac aactcgttac ttgggtcctc ccttcgatcc tttccatctc cttgttattg      780
agtgtctgct tctatttat ttatagaaga aagaccactc cttcagtcaa atatgccaaa       840
aatcatcgtc tctatgaacc accaatggaa ttagagccta tggtttatc agaagcagtc       900
tactcgacct ccttggagga agtgagtccc ttggtcaagg gagctggaaa attcaccttt      960
gatcaactta ttcaagctac cttgctagat gtgatagacc gtgggaatgt ctctatcatt     1020
tcagaaggag atgcagttga tttgaggcta gtaaaagaag atggtttgtc aagctttgag     1080
aaagactgcc taaatctagc tttttcaggt aaaaaagaag aaactctttc caatttgttt     1140
gcggattaca aggtatctga tagtctttat cgtagagcca aagttctga tgaaaaacgg      1200
attcaagcaa gagggcttca actcaaatct tcttttgaag aggtattgaa ccagatgcaa      1260
gaaggagtga aaaacgagt ttccttctgg gggctcccag attattatcg tcctttaact       1320
ggtggggaaa aggccttgca agtgggtatg ggtgccttga ctatcctgcc cctatttatc      1380
ggatttggtt tgttcttgta cagtttagac gttcatggct atctttacct cccttttgcca     1440
atacttggtt ttctagggtt agttttgtct gtttttctatt attggaagct tcgactagat     1500
aatcgtgatg gtgttctaaa tgaagcggga gctgaggtct actatctctg gaccagtttt     1560
gaaaatatgt tgcgtgagat tgcacgattg gatcaggctg aactggaaag tattgtggtc     1620
tggaatcgcc tcttggtcta tgcgacccta tttggctatg cggacaaggt tagtcatttg     1680
atgaaggttc atcagattca agtggaaaat ccagatatca atctctatgt agcttatggc      1740
tggcacagta cgttttatca ttcaacagca caaatgagcc attatgctag tgtcgcaaat     1800
acagcaagca cctactctgt atcttctgga agtggaagtt ctggtggtgg cttctctgga     1860
ggcggaggtg cggcagtat cggtgccttt                                       1890


SeqID 23
atgaaatcaa tcaacaaatt tttaacaatg cttgctgcct tattactgac agcgagtagc       60
ctgttttcag ctgcaacagt ttttgcggct gggacgacaa caacatctgt taccgttcat      120
aaactattgg caacagatgg ggatatggat aaaattgcaa atgagttaga aacaggtaac      180
tatgctggta ataaagtggg tgttctacct gcaaatgcaa aagaaattgc cggtgttatg      240
ttcgtttgga caaatactaa taatgaaatt attgatgaaa atggccaaac tctaggagtg      300
aatattgatc cacaaacatt taaactctca ggggcaatgc cggcaactgc aatgaaaaaa      360
ttaacagaag ctgaaggagc taaatttaac acggcaaatt taccagctgc taagtataaa .    420
atttatgaaa ttcacagttt atcaacttat gtcggtgaag atggagcaac cttaacaggt .    480
cctaaagcag ttccaattga aattgaatta ccattgaacg atgttgtgga tgcgcatgtg     540
catccaaaaa atacagaagc aaagccaaaa attgataaag atttcaaagg taaagcaaat     600
ccagatacac cacgtgtaga taaagataca cctgtgaacc accaagttgg agatgttgta     660
jagtacgaaa ttgttacaaa aattccagca cttgctaatt atgcaacaga aaactggagc     720
jatagaatga ctgaaggttt ggcattcaac aaaggtacag tgaaagtaac tgttgatgat     780
jttgcacttg aagcaggtga ttatgctcta acagaagtag caactggttt tgatttgaaa     840
:taacagatg ctggtttagc taaagtgaat gaccaaaacg ctgaaaaaac tgtgaaaatc     900
icttattcgg caacattgaa tgacaaagca attgtagaag taccagaatc taatgatgta     960
icatttaact atggtaataa tccagatcac gggaatactc caaagccgaa taagccaaat     1020
jaaaacggcg atttgacatt gaccaagaca tgggtgatg ctacaggtgc accaattccg     1080
ctggagctg aagcaacgtt cgatttggtt aatgctcaga ctggtaaagt tgtacaaact     1140
```

```
gtaactttga caacagacaa aaatacagtt actgttaacg gattggataa aaatacagaa    1200
tataaattcg ttgaacgtag tataaaaggg tattcagcag attatcaaga aatcactaca    1260
gctggagaaa ttgctgtcaa gaactggaaa gacgaaaatc caaaaccact tgatccaaca    1320
gagccaaaag ttgttacata tggtaaaaag tttgtcaaag ttaatgataa agataatcgt    1380
ttagctgggg cagaatttgt aattgcaaat gctgataatg ctggtcaata tttagcacgt    1440
aaagcagata aagtgagtca agaagagaag cagtggttg ttacaacaaa ggatgcttta     1500
gatagagcag ttgctgctta taacgctctt actgcacaac aacaaactca gcaagaaaaa    1560
gagaaagttg acaaagctca agctgcttat aatgctgctg tgattgctgc caacaatgca    1620
tttgaatggg tggcagataa ggacaatgaa aatgttgtga aattagtttc tgatgcacaa    1680
ggtcgctttg aaattacagg ccttcttgca ggtacatatt acttagaaga aacaaaacag    1740
cctgctggtt atgcattact aactagccgt cagaaatttg aagtcactgc aacttcttat    1800
tcagcgactg gacaaggcat tgagtatact gctggttcag gtaaagatga cgctacaaaa    1860
gtagtcaaca aaaaaatcac tatcccacaa acgggtggta ttggtacaat tatctttgct    1920
gtagcggggg ctgcgattat gggtattgca gtgtacgcat atgttaaaaa caacaaagat    1980
gaggatcaac ttgct                                                     1995
```

SeqID 24
```
atggcggtaa tggcgtatcc gctggtgtct cgcttgtatt atcgagtgga atcaaatcaa     60
caaattgctg actttgataa ggaaaaagca acgttggatg aggctgacat tgatgaacga    120
atgaaattgg cacaagcctt caatgactct ttgaataatg tagtgagtgg cgatccttgg    180
tcggaagaaa tgaagaaaaa agggcgagca gagtatgcac gtatgttaga aatccatgag    240
cggatggggc atgtggaaat ccccgttatt gacgtggatt tgccggttta tgctggtact    300
gctgaagagg tattgcagca aggggctggg catctagagg gaacttctct gccgatcgga    360
ggcaattcga cccatgcggt gattacggca catacaggtt tgccaacagc taagatgttt    420
acggatttga ccaaacttaa agttggggat aagtttatg tgcacaatat caaggaagtg     480
atggcctatc aagtggatca agtaaaggtg attgagccga cgaactttga tgatttattg    540
attgtaccag gtcatgatta tgtgaccttg ctgacttgta cgccatacat gatcaatacc    600
catcgtctat tggttcgggg gcatcggata ccgtacgtag cagaggttga ggaagaattt    660
attgcagcaa acaaactcag tcatctctat cgctacctgt tttatgtggc agttggtttg    720
attgtgattc ttttatggat tattcgacgc ttgcgcaaga agaaaaaaca accggaaaag    780
gctttgaagg cgctgaaagc agcaaggaag gaagtgaagg tggaggatgg acaacag       837
```

SeqID 25
```
atgtcaagga ctaaactacg agccttattg ggatacttgt tgatgttggt agcctgtttg     60
attcctattt attgttttg acagatggtg ttgcagtctc ttggacaggt gaaaggtcat     120
gctacatttg tgaaatccat gacaactgaa atgtaccaag aacaacagaa ccattctctc    180
gcctacaatc aacgcttggc ttcgcaaaat cgcattgtag atcctttttt ggcggaggga    240
tatgaggtca attaccaagt gtctgacgac cctgatgcag tctatggtta cttgtctatt    300
ccaagttgg aaatcatgga gccggtttat ttgggagcag attatcatca tttagggatg     360
ggcttggctc atgtggatgg tacaccgctg cctctggatg gtacagggat tcgctcagtg    420
attgctgggc accgtgcaga gccaagccat gtctttttcc gccatttgga tcagctaaaa    480
gttggagatg ctctttatta tgataatggc caggaaattg tagaatatca gatgatggac    540
acagagatta ttttaccgtc ggaatgggaa aaattagaat cggttagctc taaaaatatc    600
atgacccttga taacctgcga tccgattcct acctttaata aacgcttatt agtgaatttt    660
gaacgagtcg ctgtttatca aaaatcagat ccacaaacag ctgcagttgc gagggttgct    720
tttacgaaag aaggacaatc tgtatcgcgt gttgcaacct ctcaatggtt gtaccgtggg    780
ctagtggtac tggcatttct gggaatcctg tttgtttttgt ggaagctagc acgtttacta    840
cgagggaaa                                                            849
```

SeqID 26
```
atgaagaatc cattttttga aagacgttgt cgttacagta ttcgtaagtt atcagtagga     60
gcctgctcgc tgatgattgg tgctgtttta tttgctggtc cagccttggc tgaagaaact    120
gcagttcctg aaaatagcgg agctaataca gagcttgttt caggagagag tgagcattcg    180
accaatgaag ctgataagca gaatgaaggg gaacatgcta gagaaaacaa gctagaaaag    240
gcagaaggag tagcgatagc atctgaaact gcttcgccag caagcaatga agctgcaact    300
actgaaactg cagaagcagc tagcgcagct aaaccagagg aaaaagcaag tgaggtggtt    360
gcagaaacac catctgcaga agcaaaacct aagtctgaca aggaaacaga agcaaagccc    420
gaagcaacta accaaggga tgagtctaaa ccagcagcag aagctaataa gactgaaaaa     480
gaagtccagc cagatgtccc taaaaataca gaaaaaacat taaaaccaaa ggaaatcaaa    540
tttaattctt gggaagaatt gttaaaatgg gaaccaggtg ctcgtgaaga tgatgctatt    600
aaccgcggat ctgttgtcct cgcttcacgt cggacaggtc atttagtcaa tgaaaaagct    660
agcaaggaag caaagttca gccttatca aacaccaatt ctaaagcaaa agaccatgct     720
tctgttggtg agaagagtt caaggcctat gcttttgact attggcaata tctagattca    780
atggtcttct gggaaggtct cgtaccaact cctgacgtta ttgatgcagg tcaccgtaac    840
ggggttcctg tatacggtac actcttcttc aactggtcta atagtattgc agatcaagaa    900
agatttgctg aagctttgaa gcaagacgca gatggtagct tcccaattgc ccgtaaattg    960
gtagacatgg ccaagtatta tggctatgat ggctatttca tcaaccaaga aacaactgga    1020
gatttggtta aacctcttgg agaaaagatg cgccagttta tgctctatag caaggaatat    1080
```

```
gctgctaagg taaaccatcc aatcaagtat tcttggtacg atgccatgac ctataactat   1140
ggacgttatc atcaagatgg tttgggagaa tacaactacc aattcatgca accagaagga   1200
gataaggttc cggcagataa cttctttgct aactttaact gggataaggc taaaaatgat   1260
tacactattg caactgccaa ctggattggt cgtaatcctt atgatgtatt tgcaggtttg   1320
gaattgcaac agggtggttc ctacaagaca aaggttaagt ggaatgacat tttagacgaa   1380
aatgggaaat tgcgcctttc tcttggttta tttgccccag ataccattac aagtttagga   1440
aaaactggtg aagattatca taaaaatgaa gatatcttct ttacaggtta tcaaggagac   1500
cctactggcc aaaaaccagg tgacaaagat tggtatggta ttgctaacct agttgcggac   1560
cgtacgccag cggtaggtaa tacttttact acttctttta atacaggtca tggtaaaaaa   1620
tggttcgtag atggtaaggt ttctaaggat tctgagtgga attatcgttc agtatcaggt   1680
gttcttccaa catggcgctg tgtggcagact tcaacagggg aaaaacttcg tgcagaatat   1740
gattttacag atgcctataa tggcggaaat tcccttaaat tctctggtga tgtagccggt   1800
aagacagatc aggatgtgag actttattct actaagttag aagtaactga gaagaccaaa   1860
cttcgtgttg cccacaaggg aggaaaaggt tctaaagttt atatggcatt ctctacaact   1920
ccagactaca aattcgatga tgcagatgca tggaaagagc taacccttc tgacaactgg   1980
acaaatgaag aatttgatct tagctcacta gcgggtaaaa ccatctatgc agtcaaacta   2040
tttttcgagc atgaaggtgc tgtaaaagat tatcagttta acctaggaca attaactatc   2100
tcggacaatc accaagagcc acaatcgccg acaagctt ctgtagtgaa acaatctctt   2160
aaaaatgccc aagaagcgga agcagttgtg caatttaaag gcaacaagga tgcagatttc   2220
tatgaagttt atgaaaaaga tggagacagc tggaaattac taactggctc atcttctaca   2280
actatttatc taccaaaagt tagccgctca gcaagtgctc agggtacaac tcaagaactg   2340
aaggttgtag cagtcggtaa aaatggagtt cgttcagaag ctgcaaccac aacctttgat   2400
tggggtatga ctgtaaaaga taccagccta ccaaaaccac tagctgaaaa tatcgttcca   2460
ggtgcaacag ttattgatag tactttccct aagactgaag gtggagaagg tattgaaggt   2520
atgttgaacg gtaccattac tagcttgtca gataaatggt cttcagctca gttgagtggt   2580
agtgtggata ttcgtttgac caagccacgt accgttgtta gatgggtcat ggatcatgca   2640
ggagcgggtg gtgagtctgt taacgatggc ttgatgaaca ctaaagactt tgacctttat   2700
tataaagatg cagatggtga gtggaagcta gctaaggaag tccgtggtaa caaagcacac   2760
gtgacagata tcactcttga taaaccaatc actgctcaag actggcgctt gaatgttgtc   2820
acttctgaca atggaactcc atggaaggct attcgtatct ataactggaa aatgtatgaa   2880
aagcttgata ctgagagtgt caatattccg atggccaagg ctgcagcccg ttctctaggc   2940
aataacaagg tacaagttgg cttttgcagat gtaccggctg gagcaactat taccgtttat   3000
gataatccaa attctcaaac tccgctcgca accttgaaga gcgaagttgg aggagaccta   3060
gcaagtgcac cattggattt gacaaatcaa tctggtcttc tttattatcg tacccagttg   3120
ccaggcaagg aaattagtaa tgtcctagca gtttccgttc caaaagatga cagaagaatc   3180
aagtcagtca gcctagaaac aggacctaag aaaacaagct acgccgaagg ggaggatttg   3240
gaccttagag gtggtgttct tcgagttcag tatgaaggag gaactgagga cgaactcatt   3300
cgcctaactc acgcaggtgt atcagtatca ggttttgata cgcatcataa gggagaacag   3360
aatcttactc tccaatattt gggacaaccg gtaaatgcta atttgtcagt gactgtcact   3420
ggccaagacg aagcaagtcc gaaaactatt ttgggaattg aagtaagtca ggaaccgaaa   3480
aaagattacc tagttggtga tagcttagac ttgtctgaag gacgctttgc agtggcttat   3540
agcaatgaca ccatggaaga acattccttt actgatgagg gagttgaaat ttctggttac   3600
gatgctcaaa agactggtcg tcaaaccttg acgcttcatt accaaggcca tgaagttagc   3660
tttgatgttt tggtatctcc aaaaagcagca ttgaacgatg agtacctcaa acaaaaatta   3720
gcagaagttg aagctgctaa gaacaaggtg gtctataact ttgcttcatc agaagtaaaa   3780
gaagccttct tgaaagcaat tgaagcggcc gaacaagtgt tgaaagacca tgaaactagc   3840
acccaagatc aagtcaatga ccgacttaat aaattgacag aagctcataa agctctgaat   3900
ggtcaagaga aatttacgga agaaaagaca gagcttgatc gcttaacagg tgaggttcaa   3960
gaactcttgg ctgccaaacc aaaccatcct tcaggttctg ccctagctcc gcttcttgag   4020
aaaaacaagg ccttggttga aaaagtagat ttgagtccag aagagcttac aacagcgaaa   4080
cagagtctaa aagatctggt tgctttattg aaagaagaca agccagcagt cttttctgat   4140
agtaaaacag gtgttgaagt acacttctca aataaagaga agactgtcat caagggtttg   4200
aaagtagagc gtgttcaagc aagtgctgaa gagaagaaat actttgctgg agaagatgct   4260
catgtctttg aaatagaagg tttggataaa agtgttgatct ctcttatgct   4320
tctattgtga aaatcccaat tgaaaaagat aagaaagtta agaaagtatt tttcttacct   4380
gaaggcaaag aggcagtaga attggctttt gaacaaacgg atagtcatgt tatctttaca   4440
gcacctcact ttactcatta tgcctttgtt tatgaatctg ctgaaaaacc acaacctgct   4500
aaaccagcac cacaaaacac agtccttcca aaacctactt atcaaccgac ttctgatcaa   4560
caaaaggctc ctaaattgga agttcaagag gaaaaggttg cctttcatcg tcaagagcat   4620
gaaaatactg agatgctagt tggggaacaa cgagtcatca tacagggacg agatggactg   4680
ttaagacatg tctttgaagt tgatgaaaac ggtcagcgtc gtcttcgttc aacagaagtc   4740
atccaagaag cgattccaga aattgttgaa attggaacaa aagtaaaaac agtaccagca   4800
gtagtagcta cacaggaaaa accagctcaa aatacagcag ttaaatcaga agaagcaagc   4860
aaacaattgc caaatacagg aacagctgat gctaatgaag ccctaatagc aggcttagcc   4920
agccttggtc ttgctagttt agccttgacc ttgagacgga aaagagaaga taaagat     4977
```

SeqID 27

```
atgtcaatta catcatttgt aaaaagaatt caagatatca ctcgaaacga tgctggtgtt        60
```

```
aatggtgatg ctcaacgtat tgagcaaatg tcttggttat tattcttaaa aatttatgat     120
agccgtgaaa tggtttggga attagaagaa gacgagtatg agtcaattat cccagaggaa     180
ttaaaatggc gaaattgggc tcatgctcaa aatgggaac gggtattgac aggcgatgaa      240
ttacttgatt ttgtcaataa caagttattc aaagagttga aagagcttga ataacttca      300
aatatgccta ttcgaaaaac gattgttaaa tcagcttttg aagatgcgaa caactatatg     360
aaaaatggcg tcttgttacg ccaagtcatc aatgttattg atgaagttga tttcaatagc     420
cctgaagatc gtcattcgtt taatgatatt tacgaaaaaa ttcttaaaga tattcaaaat     480
gctgggaact caggagaatt ttatacgcca cgtgcagcga ctgattttat tgccgaagtt     540
cttgacccaa aacttggaga atcaatggca gaccttgctt gcggaacagg aggcttcttg     600
acttcgactc tgaaccgttt aagtagtcaa cgtaaaacta gtgaagatac caaaaaatat     660
aatacagctg tttttggtat tgaaaagaaa gcatttcctc atcttttagc agttacaaat     720
ctgtttcttc acgaaattga tgaccctaaa attgttcatg gaaatacttt ggagaaaaat     780
gttcgtgaat atacggatga tgaaaaattt gacattatta tgatgaatcc accttttgga     840
gggtcagaat tagaaacaat aaaaaataac tttccagcag aattacggag ttctgaaaca     900
gctgatttat ttatggctgt cattatgtat cgtttgaaag aaaatggtcg tgttggagtt     960
attttacctg atggttttct atttggtgaa ggtgtaaaac tcgcttgaa acaaaaactg     1020
gtagatgagt tcaacttgca tacgattatt aggttgcctc atagtgtctt tgcaccgtat     1080
acaggaatcc atacgacaca tcttttcttt gataaaacaa agaaaacaga agaaacttgg     1140
ttttatcgtt tagatatgcc agatggttat aaaaaatttct cgaaaactaa gccgatgaag     1200
tcagaacact tcaatcctgt tcgtgactgg tgggaaaatc gtgaagagat tctggaaggt     1260
aagttctaca aatctaaatc atttacacct agtgaattgg ctgagttgaa ttataatttta     1320
gaccagtgtg actttccaaa agaggaagag gaaatcttaa atcccttttga gttgattcag     1380
aattatcaag cggaaagagc aactttaaat cataagattg ataatgtatt agctgatatt    1440
ttgcagttgt tggaggacaa a                                              1461
```

SeqID 28
```
atgaacaata ctgaatttta tgatcgtctg ggggtatcca aaaacgcttc ggcagacgaa      60
atcaaaaagg cttatcgtaa gctttccaaa aaatatcacc cagatatcaa caaggagcct     120
ggtgctgagg acaagtacaa ggaagttcaa gaagcctatg agactttgag tgacgaccaa     180
aaacgtgctg cctatgacca gtatggtgct gcaggcgcca atggtggttt tggtggagct     240
ggtggtttcg gcggtttcaa tggggcaggt ggcttcggtg gttttgagga tattttctca     300
agtttcttcg gcggaggcgg ttcttcgcgc aatccaaacg ctcctcgcca aggagatgat     360
ctccagtatc gtgtcaattt gacctttgaa gaagctatct tcggaactga gaaggaagtt     420
aagtatcatc gtgaagctgg ctgtcgtaca tgtaatggat ctggtgctaa gccagggaca     480
agtccagtca cttgtggacg ctgtcatgga gctgggtgtca ttaacgtcga tacgcagact     540
cctcttggta tgatgcgtcg ccaagtaacc tgtgatgtct gtcacggtcg aggaaaagaa     600
atcaaatatc catgtacaac ctgtcatgga acaggtcatg agaaacaagc tcatagcgta     660
catgtgaaaa tccctgctgg tgtggaaaca ggtcaacaaa ttcgcctcgc tggtcaaggt     720
gaagcaggct ttaacggtgg accttatggt gacttgtatg tagtagtttc tgtggaagct     780
agcgacaagt ttgaacgtga aggaacgact atcttctaca atctcaacct caactttgtc     840
caagcggctc ttggtgatac agtagatatt ccaactgttc acggtgatgt tgaattggtt     900
attccagagg gaactcagac tggtaagaag ttccgcctac gtagtaaggg ggcaccgagc     960
cttcgtggcg gtgcagttgg tgaccaatac gttactgtta atgtcgtaac accgacaggc     1020
ttgaacgacc gccaaaaagt agccttgaaa gaattcgcgg ctgctggtga cttgaaagta     1080
aatccaaaga aaaaggctt cttttgaccat attaaagatg cctttgatgg agaa          1134
```

SeqID 29
```
atgaatccta atctttttag aagcgtcgag ttttatcaga gacgttacca taactatgcg      60
acagtgttaa ttatacctct ttcattacta tttactttca tcttgatttt ctcccttgtt     120
gccacaaaag aaattactgt tacttcccaa ggagaaatcg ccctacaag tgtcattgcc       180
tccattcagt caaccagtga taatcctatc ctagctaatc atttagtggc aaatcaagta     240
gttgaaaaag gggacttact catcaaatac tctgaagaaca tggaagaaac tcagaaaact     300
gccttagcaa ctcaattaca aagacttgag aagcaaaaag aaggacttgg aattttgaaa     360
caaagcttag aaaaagcgac tgatctttt tctggcgagg atgaatttgg ctaccataat     420
accttttatga attttactaa acaatcccat gatattgaac tgggtatcac aaagactaac     480
accgaagttt caaatcaagc taatctttcc aatagcagtt catcagctat tgaacaagaa     540
attacaaaag ttcaacaaca aattggagaa tatcaagagt tgagagatgc tatcataaat     600
aacagagcac gcttaccaac tggcaatccg caccagtcaa ttttgaatcg ttatcttgta     660
gcctcacaag acaaacaca aggaactgca gaggagccat ttttatctca aattaatcaa      720
agtattgcag gtcttgaatc atctatcgca agcctcaaaa ttcagcaagc tggtatcgga    780
agtgtagcaa cttatgataa cagtttatga accaaaattg aagtactccg cactcagttt     840
ttacagacag cctcacagta acaactaact gtggagaatc aattaacaga attaaaagta     900
caactagatc aagccacaca gcgtttggaa aacaatacct taacctcccc aagtaaaggt     960
atcgttcatc tgaacagcga atttgaaggt aaaaatagaa ttccaactgg tacagaaatt     1020
gctcaaatat tccctgtcat cacagataca agagaagtac taatcactta ctacgtatct     1080
tctgactatc tacctctact agataaagga caaactgtaa gattaaaact ggagaagatt     1140
ggaaatcacg gcaccaccat catcggccaa cttcagacaa ttgatcaaac tcctaccaga     1200
acagagcaag gaaatctctt taaattaacc gctcttgcaa aactatctaa cgaggatagt     1260
```

```
aaactcatcc aatatggctt acaaggtcgc gtcactagtg taactacaaa gaaaacatat    1320
tttgattatt tcaaagataa aattttaaca cattctgat                          1359


SeqID 30
atgtcaaaga aactcaatcg taaaaaacaa ttacgaaatg gcctccgtcg cgcaggtgcc      60
ttttcaagta cggtgctgacta ggttgtagat gagacaaaaa aagtcgtgaa gcgtgcagaa   120
cagtcagcaa gcgcagctgg taaggctgtt tctaaaaaag ttgaacaagc agtagaagct     180
accaaagagc aagctcaaaa agtagctaat tctgtagaag attttgcagc aaatttgggt     240
ggacttccac ttgatcgtgc caagactttc tatgatgaag gaatcaagtc tgcttcagat     300
ttcaaaaact ggactgaaaa agaactcctt gccttgaaag gaatcggccc agctaccatc     360
aagaaattga aagaaatgg catcaagttc aag                                   393


SeqID 31
ttgattagcc ttttcggcct tgctgctgcc aaaccagtcc aggctgatac aagtatcgca     .60
gacattcaaa aaagaggcga actggttgtc ggtgtcaaac aagacgttcc caatttttggt   120
tacaaagatc ccaagaccgg tacttattct ggtatcgaaa ccgacttggc caagatggta     180
gctgatgaac tcaaggtcaa gattcgctat gtgccggtta cagcacaaac ccgcggcccc    .240
cttctagaca atgaacaggt cgatatggat atcgcgacct ttaccatcac ggacgaacgc     300
aaaaaactct acaactttac cagtccctac tacacagacg cttctggatt tttggtcaat     360
aaatctgcca aaatcaaaaa gattgaggac ctaaacggca aaaccatcgg agtcgcccaa     420 .
ggttctatca cccaacgcct gattactgaa ctgggtaaaa agaaaggtct gaagtttaaa     480
ttcgtcgaac ttggttccta cccagaattg attacttccc tgcacgctca tcgtatcgat     540
accttttccg ttgaccgctc tattctatct ggctacacta gtaaacggac agcactacta     600
gatgatagtt tcaagccatc tgactacggt attgttacca agaaatcaaa tacagagctc     660
aacgactatc ttgataactt ggttactaaa tggagcaagg atggtagttt gcagaaactt     720
tatgaccgtt acaagctcaa accatctagc catactgcag at                       762


SeqID 32
atgagtaata tcagtttaac aacacttggt ggtgtgcgtg agaatggaaa aaatatgtac      60
attgctgaaa ttggagagtc cattttttgtt ttgaatgtag ggttaaaata tcctgaaaat   120
gaacaattag gggtcgatgt ggtgattcca aacatggatt acctttttga aaatagcgac     180
cgtattgctg gggtttttctt gacccacggg catgcggatg ccattggtgc tctaccgtat   240
ctcttggcag aggctaaagt tcctgtattt gggtctgaag tgaccattga gttggcaaag     300
ctctttgtca aaggaaatga tgccgttaag aaatttaatg atttccatgt cattgatgag     360
aatacggaga ttgattttgg tgggacagtg gtttccttct tccctacgac ttactccgtt     420
ccagagagtc tgggaattgt cttgaagaca tcggaaggaa gcatcgttta tacaggtgac     480
ttcaaatttg accaaacggc tagtgaatct tatgcaactg attttgctcg tttggcagag     540
attggtcgtg acggcgtcct ggctctcctc agtgattcgg ccaatgcaga cagcaatatt     600
caggtggcta gtgaaagtga agttagggat gaaattaccc aaactattgc tgactgggaa     660
ggtcgtatca tcgttgcagc tgtttccagt aatctttctc gtattcagca gattttttgac   720
gctgcggata aaacaggtcg acgtatcgtc ttgacaggat ttgatattga aaatatcgtc     780
cgcacagcga ttcgtcttaa gaagttgtct ttagccaacg aaattctttt gattaagcct     840
aaagatatgt ctcgctttga agaccatgag ttgattattc ttgagacagg tcgtatgggt     900
gagcctatca atggacttcg taagatgtcg attggtcgcc atcgttatgt agaaatcaag     960
gatggggacc tagtctatat tgctacggct ccgtctattg ctaaagaagc ctttgttgcg    1020
cgtgtggaaa atatgatttta tcaggcaggt ggggttgtca aattgattac ccaaagttta    1080
catgtatcag ggcacggaaa tgtgcgtgat ttgcagctga tgatcaatct tttgcaacct    1140
aagtacctct tccctgtcca aggggagtat cgtgagttgg atgctcacgc taaggctgcc    1200
atggcagttg ggatgttgcc agaacgcatc ttcattccta aaaaggggac gaccatggct    1260
tacgagaatg gagactttgt tccagctgga tcggtttcag caggagataat cttgattgat   1320
gggaatgcca ttggtgatgt tggaaatgtt gttcttcgtg accgtaaggt cttgtcagag    1380
gatggaattt tcatcgtggc tattacagtc aaccgtcgtg agaagaaaat tgtggctagg    1440
gctcgtgttc acacgcgtgg atttgtttat ctcaagaaga gtcgcgatat tctccgtgaa    1500
agttcagaat tgattaacca aacggtagaa gagtatcttc aaggagatga ctttgactgg    1560
gcagatctca aaggtaaggt tcgtgacaat ctgaccaagt acctctttga tcaaaccaag    1620
cgtcgcccag ccatttttacc agtagtcatg gaagcaaaa                          1659


SeqID 33
atgaaaaaaa gtacagtatt gtcattaacc acagctgcag ttattttagc agcctatgcc      60
cctaatgagg tagtcttagc agacacatct agctctgaag atgctttaaa catctctgat     120
aaagaaaaag tagcagaaaa taaagagaaa catgaaaata tccatagtgc tatggaaact     180
tcacaggatt ttaaagagaa gaaaacagca gtcattaagg aaaaagaagt tgttagtaaa     240
aatcctgtga tagacaataa cactagcaat gaagaagcaa aaatcaaaga agaaaattcc     300
aataaatccc aaggagatta tacggactca tttgtgaata aaaacacaga aaatcccaaa     360
aagaagata aagttgtcta tattgctgaa tttaaagata aagaatctgg agaaaaagca     420
atcaaggaac tatccagtct taagaataca aaagtttttat atacttatga tagaatttttt    480
aacggtagtg ccatagaaac aactccagat aacttggaca aaattaaaca aatagaaggt     540
atttcatcgg ttgaagggc acaaaaagtc caacccatga tgaatcatgc cagaaaggaa      600
```

```
attggagttg aggaagctat tgattaccta aagtctatca atgctccgtt tgggaaaaat    660
tttgatggta gaggtatggt catttcaaat atcgatactg gaacagatta tagacataag    720
gctatgagaa tcgatgatga tgccaaagcc tcaatgagat ttaaaaaaga agacttaaaa    780
ggcactgata aaaattattg gttgagtgat aaaatccctc atgcgttcaa ttattataat    840
ggtggcaaaa tcactgtaga aaaatatgat gatggaaggg attattttga cccacatggg    900
atgcatattg cagggattct tgctggaaat gatactgaac aagacatcaa aaactttaac    960
ggcatagatg gaattgcacc taatgcacaa attttctctt acaaaatgta ttctgacgca   1020
ggatctgggt ttgcgggtga tgaaacaatg tttcatgcta ttgaagattc tatcaaacac   1080
aacgttgatg ttgtttcggt atcatctggt tttacaggaa caggtcttgt aggtgagaaa   1140
tattggcaag ctattcgggc attaagaaaa gcaggcattc caatggttgt cgctacgggt   1200
aactatgcga cttctgcttc aagttcttca tgggatttag tagcaaataa tcatctgaaa   1260
atgaccgaca ctggaaatgt aacacgaact gcagcacatg aagatgcgat agcggtcgct   1320
tctgctaaaa atcaaacagt tgagtttgat aaagttaaca taggtggaga aagttttaaa   1380
tacagaaata taggggcctt tttcgataag agtaaaatca caacaaatga agatggaaca   1440
aaagctccta gtaaattaaa atttgtatat ataggcaagg ggcaagacca agatttgata   1500
ggtttggatc ttaggggcaa aattgcagta atggatagaa tttatacaaa ggatttaaaa   1560
aatgctttta aaaaagctat ggataaggg gcacgcgaca ttatggttgt aaatactgta   1620
aattactaca atagagataa ttggacagag cttccagcta tgggatatga agcggatgaa   1680
ggtactaaaa gtcaagtgtt ttcaattttca ggagatgatg gtgtaaagct atggaacatg   1740
attaatcctg ataaaaaaac tgaagtcaaa agaaataata aagaagattt taaagataaa   1800
ttggagcaat actatccaat tgatatggaa agttttaatt ccaacaaacc gaatgtaggt   1860
gacgaaaaag agattgactt taagtttgca cctgacacag acaaagaact ctataagaa    1920
gatatcatcg ttccagcagg atctacatct tggggggccaa gaatagattt acttttaaaa   1980
cccgatgttt cagcacctgg taaaaatatt aaatccacgc ttaatgttat taatggcaaa   2040
tcaacttatg gctatatgtc aggaactagt atggcgactc caatcgtggc agcttctact   2100
gttttgatta gaccgaaatt aaaggaaatg cttgaaagac ctgtattgaa aaatcttaag   2160
ggagatgaca aaatagatct tacaagtctt acaaaaattg ccctacaaaa tactgcgcga   2220
cctatgatgg atgcaacttc ttggaaagaa aaaagtcaat actttgcatc acctagacaa   2280
caggagcag gcctaattaa tgtggccaat gctttgagaa atgaagttgt agcaactttc   2340
aaaaacactg attctaaagg tttggtaaac tcatatggtt ccatttctct taaagaaata   2400
aaaggtgata aaaaatactt tacaatcaag cttcacaata catcaaacag accttttgact   2460
tttaaagttt cagcatcagc gataactaca gattctctaa ctgacagatt aaaacttgat   2520
gaaacatata aagatgaaaa atctccagat ggtaagcaaa ttgttccaga aattcaccca   2580
gaaaagtca aaggagcaaa tatcacattt gagcatgata ctttcactat aggcgcaaat   2640
tctagctttg atttgaatgc ggttataaat gttggagagg ccaaaaacaa aaataaattt  ⋄ 2700
gtagaatcat ttattcattt tgagtcagtg gaagaaatgg aagctctaaa ctccaacggg   2760
aagaaaataa acttccaacc ttctttgtcg atgcctctaa tgggatttgc tgggaattgg   2820
aaccacgaac caatccttga taaatgggct tgggaagaag ggtcaagatc aaaaacactg   2880
ggaggttatg atgatgatgg taaaccgaaa attccaggaa ccttaaataa gggaattggt   2940
ggagaacatg gtatagataa atttaatcca gcaggagtta tacaaaatag aaaagataaa   3000
aatacaacat ccctggatca aaatccagaa ttatttgctt tcaataacga agggatcaac   3060
gctccatcat caagtggttc taagattgct aacatttatc ctttagattc aaatggaaat   3120
cctcaagatg ctcaacttga aagaggatta acacctctc cacttgtatt aagaagtgca   3180
gaagaaggat tgatttcaat agtaaataca aataaagagg gagaaaatca aagagactta   3240
aaagtcattt cgagagaaca ctttattaga ggaattttaa attctaaaag caatgatgca   3300
aagggaatca aatcatctaa actaaaagtt tggggtgact tgaagtggga tggactcatc   3360
tataatccta gaggtagaga agaaaatgca ccagaaagta aggataatca agatcctgct   3420
actaagataa gaggtcaatt tgaaccgatt gcggaaggtc aatatttcta taaatttaaa   3480
tatagattaa ctaaagatta cccatggcag gtttcctata ttcctgtaaa aattgataac  · 3540
accgcccta agattgtttc ggttgatttt tcaaatcctg aaaaaattaa gttgattaca   3600
aaggatactt atcataaggt aaaaagatcag tataagaatg aaacgctatt tgcgagagat   3660
caaaaagaac atccctgaaca atttgacgag attgcgaacg aagtttggta tgctggcgcc   3720
gctcttgtta atgaagatgg agaggttgaa aaaaatcttg aagtaactta cgcaggtgag   3780
ggtcaaggaa gaaatagaaa acttgataaa gacggaaata ccatttatga aattaaaggt   3840
gcgggagatt taaggggaaa aatcattgaa gtcattgcat tagatggttc tagcaatttc  3900.
acaaagatc atagaattaa atttgctaat caggctgatg aaaaggggat gatttcctat   3960
tatctagtag atcctgatca agattcatct aaatatcaaa agcttggcga gattgcagaa  · 4020
tctaaattta aaaatttagg aaatgggaaa gagggtagtc taaaaaaaga tacaactggg   4080
gtagaacatc atcatcaaga aaatgaagag tctattaaag aaaaatctag tttactatt   4140
gatagaaata tttcaacaat tagagacttt gaaaataaag acttaaagaa actcattaaa   4200
aagaaattta gagaagttga tgatttaca agtaagaact gtaagagaat ggaggaatac   4260
gattataaat acgatgataa aggaaatata atagcctacg atgatgggac tgatctagaa   4320
tatgaaactg agaaacttga cgaaatcaaa tcaaaaattt atggtgttct aagtccgtct   4380
aaagatggac actttgaaat tcttggaaag ataagtaatg tttctaaaaa tgccaaggta   4440
tattatggga ataactataa atctatagaa atcaaagcga ccaagtatga tttccactca   4500
aaaacgatga catttgatct atacgctaat attaatgata ttgtggatgg attagctttt   4560
gcaggagata tgagattatt tgttaaagat aatgatcaga aaaagctga aattaaaatt   4620
agaatgcctg aaaaaattaa ggaaactaaa tcagaatatc cctatgtatc aagttatggg   4680
```

```
aatgtcatag aattagggga aggagatctt tcaaaaaaca aaccagacaa tttaactaaa   4740
atggaatctg gtaaaatcta ttctgattca gaaaaacaac aatatctgtt aaaggataat   4800
atcattctaa gaaaaggcta tgcactaaaa gtgactacct ataatcctgg aaaaacggat   4860
atgttagaag gaaatggagt ctatagcaag gaagatatag caaaaataca aaaggccaat   4920
cctaatctaa gagccctttc agaaacaaca atttatgctg atagtagaaa tgttgaagat   4980
ggaagaagta cccaatctgt attaatgtcg gctttgaacg gcttaacat tataaggtat    5040
caagtgttta catttaaaat gaacgataaa ggggaagcta tcgataaaga cggaaatctt    5100
gtgacagatt cttctaaact tgtattattt ggtaaggatg ataaagaata cactggagag   5160
gataagttca atgtagaagc tataaaagaa gatggctcca tgttatttat tgataccaaa   5220
ccagtaaacc tttcaatgga taagaactac tttaatccat ctaaatctaa taaaatttat   5280
gtacgaaatc cagaatttta tttaagaggt aagatttctg ataagggtgg ttttaactgg   5340
gaattgagag ttaatgaatc ggttgtagat aattatttaa tctacggaga tttacacatt   5400
gataacacta gagatttaa tattaagctg aatgttaaag acggtgacat catggactgg    5460
ggaatgaaag actataaagc aaacggattt ccagataagg taacagatat ggatggaaat   5520
gtttatcttc aaactggcta tagcgatttg aatgctaaag cagttggagt ccactatcag   5580
tttttatatg ataatgttaa acccgaagta aacattgatc ctaaggtgaaa tactagtatc  5640
gaatatgctg atggaaaatc tgtagtcttt aacatcaatg ataaaagaaa taatggattc   5700
gatggtgaga ttcaagaaca acatatttat ataaatggaa aagaatatac atcatttaat   5760
gatattaaac aaataataga caagacacta aacattaaga ttgttgtaaa agatttgca    5820
agaaatacaa ccgtaaaaga attcatttta aataaagata cgggagaggt aagtgaatta   5880
aaacctcata gggtaactgt gaccattcaa aatgaaaag aaatgagttc aacgatagtg    5940
tcggaagaag attttatttt acctgtttat aagggtgaat tagaaaaagg ataccaattt   6000
gatggttggg aaatttctgg tttcgaaggt aaaaaagacg ctggctatgt tattaatcta   6060
tcaaaagata cctttataaa acctgtattc aagaaaatag aggagaaaaa gggaggaagaa   6120
aataaaccta cttttgatgt atcgaaaaag aaagataacc cacaagtaaa ccatagtcaa   6180
ttaaatgaaa gtcacagaaa agaggattta caaagagaag agcattcaca aaaatctgat    6240
tcaactaagg atgttacagc tacagttctt gataaaaaca atatcagtag taaatcaact    6300
actaacaatc ctaataagtt gccaaaaact ggaacagcaa gcggagccca gacactatta   6360
gctgccggaa taatgtttat agtaggaatt tttcttggat tgaagaaaaa aaatcaagat   6420
```

SeqID 34

```
atggggaaag gccattggaa tcggaaaaga gtttatagca ttcgtaagtt tgctgtggga     60
gcttgctcag taatgattgg gacttgtgca gtttttattag gaggaaatat agctggagaa   120
tctgtagttt atgcggatga aacacttatt actcatactg ctgagaaacc taaagaggaa   180
aaaatgatag tagaagaaaa ggctgataaa gctttggaaa ctaaaaatat agttgaaagg    240
acagaacaaa gtgaacctag ttcaactgag gctattgcat ctgagaagaa agaagatgaa    300
gccgtaactc caaaagagga aaaagtgtct gctaaaccgg aagaaaaagc tccaaggata    360
gaatcacaag cttcaaatca agaaaaaccg ctcaaggaag atgctaaagc tgtaacaaat    420
gaagaagtga atcaaatgat tgaagacagg aaagtggatt ttaatcaaaa ttggtacttt    480
aaactcaatg caaattctaa ggaagccatt aaacctgatg cagacgtatc tacgtggaaa    540
aaattagatt taccgtatga ctggagtatc tttaacgatt tcgatcatga atctcctgca    600
caaatgaag gtggacagct caacggtggg gaagcttgt atcgcaagac tttcaaacta     660
gatgaaaaag acctcaagaa aaatgttcgc cttactttg atggcgtcta catggattct    720
caagtttatg tcaatggtca gttagtgggg cattatccaa atggttataa ccagttctca    780
tatgatatca ccaaataccт tcaaaaagat ggtcgtgaga atgtgattgc tgtccatgca   840
gtcaacaaac agccaagtag ccgttggtat tcaggaagtg gtatctatcg tgatgtgact   900
ttacaagtga cagataaggt gcatgttgag aaaaatggga caactatttt aacaccaaaa   960
cttgaagaac aacaacatgg caaggttgaa actcatgtga ccagcaaaat cgtcaatacg   1020
gacgacaaag accatgaact tgtagccgaa tatcaaatcg ttgaacgagg tggtcatgct   1080
gtaacaggct tagttcgtac agcgagtcgt accttaaaag cacatgaatc aacaagccta   1140
gatgcgattt tagaagttga aagaccaaga ctctggaatg ttttaaatga caaacctgcc   1200
ttgtacgaat tgattacgcg tgtttaccgt gacggtcaat tggttgatgc taagaaggat   1260
ttgtttggtt accgttacta tcactggact ccaaatgaag gtttctcttt gaatggtgaa   1320
cgtattaaat tccatggagt atccttgcac cacgaccatg gggcgcttgg agcagaagaa   1380
aactataaag cagaatatcg ccgtctcaaa caaatgaagg agatgggagt taactccatc   1440
cgtacaaccc acaaccctgc tagtgagcaa accttgcaaa tcgcagcaga actaggttta   1500
ctcgttcagg aagaggcctt tgatacgtgg tatggtggca agaaaccttta tgactatgga   1560
cgtttctttg aaaaagatgc cactcaccca gaagctcgaa aaggtgaaaa atggtctgat   1620
tttgacctac gtaccatggt cgaaagaggc aaaaacaacc ctgctatctt catgtggtca   1680
attggtaatg aaataggtga agctaatgat gatgccacct ctttagcaac tgttaaacgt   1740
ttggttaagg ttatcaagga tgttgataag actcgctatg ttaccatggg agcagataaa   1800
ttccgtttcg gtaatggtag cggagggcat gagaaaattg ctgatgaact cgatgctgtt   1860
ggatttaact attctgaaga taattacaaa gcccttagag ctaagcatcc aaaatggttg   1920
atttatggat cagaaacatc ttcagctacc cgtacacgtg gaagttacta tcgccctgaa   1980
cgtgaattga acatagcaa tggacctgag cgtaattatg aacagtcaga ttatggaaat    2040
gatcgtgtgg gttgggggaa aacagcaacc gcttcatgga cttttgaccg tgacaacgct   2100
ggctatgctg acagtttat ctggacaggt acggactata ttggtgaacc tacaccatgg    2160
cacaaccaaa atcaaactcc tgttaagagc tcttactttg gtatcgtaga tacagccggc   2220
```

```
attccaaaac atgacttcta tctctaccaa agccaatggg tttctgttaa gaagaaaccg    2280
atggtacacc ttcttcctca ctggaactgg gaaaacaaag aattagcatc caaagtagct    2340
gactcagaag gtaagattcc agttcgtgct tattcgaatg cttctagtgt agaattgttc    2400
ttgaatggaa aatctcttgg tcttaagact ttcaataaaa aacaaaccag cgatgggcgg    2460
acttaccaag aaggtgcaaa tgctaatgaa ctttatcttg aatggaaagt tgcctatcaa    2520
ccaggtacct tggaagcaat tgctcgtgat gaatctggca aggaaattgc tcgagataag    2580
attacgactg ctggtaagcc agcggcagtc cgtcttatta aggaagacca tgcgattgca    2640
gcagatggaa aagacttgac ttacatctac tatgaaattg ttgacagcca ggggaatgtg    2700
gttccaactg ctaataatct ggttcgcttc caattgcatg gccaaggtca actggtcggt    2760
gtagataacg gagaacaagc cagccgtgaa cgctataagg cgcaagcaga tggttcttgg    2820
attcgtaaag catttaatgg taaaggtgtt gccattgtca aatcaactga acaagcaggg    2880
aaattcaccc tgactgccca ctctgatctc ttgaaatcga accaagtcac tgtctttact    2940
ggtaagaaag aaggacaaga gaagactgtt ttggggacag aagtgccaaa agtacagacc    3000
attattggag aggcacctga aatgcctacc actgttccgt ttgtatacag tgatggtagc    3060
cgtgcagaac gtcctgtaac ctggtcttca gtagatgtga gcaagcctgg tattgtaacg    3120
gtgaaaggta tggctgacgg acgagaagta gaagctcgtg tagaagtgat tgctcttaaa    3180
tcagagctac cagttgtgaa acgtattgct ccaaatactg acttgaattc tgtagacaaa    3240
tctgtttcct atgtttttgat tgatggaagt gttgaagagt atgaagtgga caagtgggag    3300
attgccgaag aagataaagc taagttagca attccaggtt ctcgtattca agcgaccggt    3360
tatttagaag gtcaaccaat tcatgcaacc cttgtggtag aagaaggcaa tcctgcggca    3420
cctgcagtac caactgtaac ggttggtggt gaggcagtaa caggtcttac tagtcaaaaa    3480
ccaatgcaat accgcactct tgcttatgga gctaagttgc cagaagtcac agcaagtgct    3540
aaaaatgcag ctgttacagt tcttcaagca agcgcagcaa acggcatgcg tgcgagcatc    3600
tttattcagc ctaaagatgg tggccctctt caaacctatg caattcaatt ccttgaagaa    3660
gcgccaaaaa ttgctcactt gagcttgcaa gtggaaaaag ctgacagtct caaagaagac    3720
caaactgtca aattgtcggt tcgagctcac tatcaagatg gaacgcaagc tgtattacca    3780
gctgataaag taaccttctc tacaagtggt gaaggggaag tcgcaattcg taaaggaatg    3840
cttgagttgc ataagccagg agcagtcact ctgaacgctg aatatgaggg agctaaagac    3900
caagttgaac tcactatcca agccaatact gagaagaaga ttgcgcaatc catccgtcct    3960
gtaaatgtag tgacagattt gcatcaggaa ccaagtcttc cagcaacagt aacagttgag    4020
tatgacaaag gtttccctaa aactcataaa gtcacttggc aagctattcc gaaagaaaaa    4080
ctagactcct atcaaacatt tgaagtacta ggtaaagttg aaggaattga ccttgaagcg    4140
cgtgcaaaag tctctgtaga aggtatcgtt tcagttgaag aagtcagtgt gacaactcca    4200
atcgcagaag caccacaatt accagaaagt gttcggacat atgattcaa tggtcacgtt    4260
tcatcagcta aggttgcatg ggatgcgatt cgtccagagc aatacgctaa ggaaggtgtc    4320
tttacagtta atggtcgctt agaaggtacg caattaacaa ctaaacttca tgttcgcgta    4380
tctgctcaaa ctgagcaagg tgcaaacatt tctgaccaat ggaccggttc agaattgcca    4440
cttgcctttg cttcagactc aaatccaagc gacccagttt caaatgttaa tgacaagctc    4500
atttcctaca ataaccaacc agccaatcgt tggacaaact ggaatcgtac taatccagaa    4560
gcttcagtcg gtgttctgtt tggagattca ggtatcttga gcaaacgctc cgttgataat    4620
ctaagtgtcg gattccatga agaccatgga gttggtgtac cgaagtctta tgtgattgag    4680
tattatgttg gtaagactgt cccaacagct cctaaaaacc ctagttttgt tggtaatgag    4740
gaccatgtct ttaatgattc tgccaactgg aaaccagtta ctaatctaaa agccctgct    4800
caactcaagg ctggagaaat gaaccacttt agctttgata aagttgaaac ctatgctgtt    4860
cgtattcgca tggttaaagc agataacaag cgtggaacgt ctatcacaga ggtacaaatc    4920
tttgcgaaac aagttgcggc agccaagcaa ggacaaacaa gaatccaagt tgacggcaaa    4980
gacttagcaa acttcaaccc tgatttgaca gactactacc ttgagtctgt agatggaaaa    5040
gttccggcag tcacagcaag tgttagcaac aatggtctcg ctaccgtcgt tccaagcgtt    5100
cgtgaaggtg agccagttcg tgtcatcgcg aaagctgaaa atggcgacat cttaggagaa    5160
taccgtctgc acttcactaca ggataagagc ttactttctc ataaaccagt tgctgcggtt    5220
aaacaagctc gcttgctaca agtaggtcaa gcacttgaat tgccgactaa ggttccagtt    5280
tacttcacag gtaaagacgg ctacgaaaca aaagacctga cagttgaatg ggaagaagtt    5340
ccagcggaaa atctgacaaa agcaggtcaa tttactgttc gaggccgtgt ccttggtagt    5400
aaccttgttg ctgagatcac tgtacgagtg acagacaaac ttggtgagac tctttcagat    5460
aaccctaact atgatgaaaa cagtaaccag gcctttgctt cagcaaccaa tgatattgac    5520
aaaactctc atgaccgcgt tgactatctc aatgacggag atcattcaga aaatcgtcgt    5580
tggacaaact ggtcaccaac accatcttct aatccagaag tatcagcggg tgtgattttc    5640
cgtgaaaatg gtaagattgt agaacggact gttacacaag aaaagttca gttctttgca    5700
gatagtggta cggatgcacc atctaaactc gttttagaac gctatgtcgg tccagagttt    5760
gaagtgccaa cctactattc aaactaccaa gcctacgacg cagaccatcc attcaacaat    5820
ccagaaaatt gggaagctgt tccttatcgt gcggataaag acattcagc tggtcatgaa    5880
atcaacgtaa catttaaagc tatcaaagcc aaagctatga gatggcgtat ggagcgtaaa    5940
gcagataaga gcggtgttgc gatgattgag atgaccttcc ttgcaccaag tgaattgcct    6000
caagaaagca ctcaatcaaa gattcttgta gatggaaaag aacttgctga tttcgctgaa    6060
aatcgtcaag actatcaaat tacctataaa ggtcaacggc aaaagtctc agttgaagaa    6120
acaatcaag tagcttcaac tgtggtagat agtggagaag atagctttcc agtacttgtt    6180
cgcctcgttt cagaaagtgg aaaacaagtc aaggaatacc gtatccactt gactaaggaa    6240
aaaccagttt ctgagaagac agttgctgct gtacaagaag atcttccaaa aatcgaattt    6300
```

```
gttgaaaaag atttggcata caagacagtt gagaaaaaag attcaacact gtatctaggt   6360
gaaactcgtg tagaacaaga aggaaaagtt ggaaaagaac gtatctttac agcgattaat   6420
cctgatggaa gtaaggaaga aaaactccgt gaagtggtag aagttccgac agaccgcatc   6480
gtcttggttg gaaccaaacc agtagctcaa gaagctaaaa aaccacaagt gtcagaaaaa   6540
gcagatacaa aaccaattga ttcaagtgaa gctagtcaaa ctaataaagc ccagttacca   6600
agtacaggta gtgcggcaag ccaagcagca gtagcagcag gtttaactct tctaggtttg   6660
agtgcaggat tagtagttac taaaggtaaa aaagaagac                          6699


SeqID 35
atggctcctt ctgtagtgga cgcagccacc tatcactatg taaataaaga gattatttca     60
caagaagcta aagatttaat tcagacagga aagcctgaca ggaatgaagt tgtatatggt    120
ttggtgtatc aaaaagatca gttgcctcaa acaggacag aagcatctgt tttgacagct    180
tttggtttgc tgactgttgg gagcttgctt ttaatctaca agagaaagaa aattgctagc    240
gtctttctag ttggagctat gggattggta gttcttccta gtgcaggggc tgtagaccca    300
gttgcgaccc tagcgctggc tagtcgagag ggtgttgttg aaatggaggg ctatcgctat    360
gttggttatc tatcaggtga catcctcaaa acgcttggct tggacactgt tttagaagaa    420
acctcagcaa aacctggaga ggtgactgtg gtcgaagttg agactcctca atcaataaca    480
aatcaggagc aagctaggac agaaaaccaa gtagtagaga cagaggaagc tccaaaagaa    540
gaagcaccta aaacagaaga aagtccaaag gaagaaccaa aatcggaggt aaaacctact    600
gacgcaccc ttcctaaagt agaagagggg aaagaagatt cagcagaacc agctccagtt    660
gaagaagtag gtggagaagt tgagtcaaaa ccagaggaaa aagtagcagt taagccagaa    720
agtcaaccat cagacaaacc agctgaggaa tcaaaagttg aacaagcagg tgaaccagtc    780
gcgccaagag aagacgaaaa ggcaccagtc gagccagaaa agcaaccaga agctcctgaa    840
gaagagaagg ctgtagagga aacaccgaaa caagaagagt caactccaga taccaaggct    900
gaagaaactg tagaaccaaa agaggagact gttaatcaat ctattgaaca accaaaagtt    960
gaaacgcctg ctgtagaaaa acaaacagaa ccaacagagg aaccaaaagt tgaacaagca   1020
ggtgaaccag tcgcgccaag agaagacgaa caggcaccaa cggcaccagt tgagccagaa   1080
aagcaaccag aagttcctga agaagagaag gctgtagagg aaacaccgaa accagaagat   1140
aaaataaagg gtattggtac taaagaacca gttgataaaa gtgagttaaa taatcaaatt   1200
gataaagcta gttcagtttc tcctactgat tattctacag caagttacaa tgctcttgga   1260
cctgttttag aaactgcaaa aggtgtctat gcttcagagc ctgtaaaaca gcctgaggta   1320
aatagcgaga caaataaact taaaacggct attgacgctc taaacgttga taaaactgaa   1380
ttaaacaata cgattgcaga tgcaaaaaca aaggtaaaag aacattacag tgatagaagt   1440
tggcaaaacc tccaaactga agttacaaag gctgaaaaag ttgcagctaa tacagatgct   1500
aaacaaagtg aagttaacga agctgttgaa aaattaactg caactattga aaaattggtt   1560
gaattatctg aaaagccaat attaacattg actagtaccg ataagaaaat attggaacgt   1620
gaagctgttg ctaagtatac tctagaaaat caaaacaaaa caaaaatcaa atcaatcaca   1680
gctgaattga aaaaaggaga agaagttatt aatactgtag tccttacaga tgacaaggta   1740
acaacagaaa ctataagcgc tgcatttaag aacctagagt actacaaaga atacacccta   1800
tctacaacta tgatttacga cagaggtaac ggtgaagaaa ctgaaactct agaaaatcaa   1860
aatattcaat tagatcttaa aaaagttgag cttaaaaata ttaaacgtac agatttaatc   1920
aaatacgaaa atggaaaaga aactaatgaa tcactgataa caactattcc tgatgataag   1980
agcaattatt atttaaaaat aacttcaaat aatcagaaaa ctacattact agctgttaaa   2040
aatatagaag aaactacggt taacggaaca cctgtatata aagttacagc aatcgcagac   2100
aatttagtct ctagaactgc tgataataaa tttgaagaag aatacgttca ctatattgaa   2160
aaacctaaag tccacgaaga taatgtatat tataatttca agaattagt ggaagctatt   2220
caaaacgatc cttcaaaaga atatcgtctg ggacaatcaa tgagcgctag aaatgttgtt   2280
cctaatggaa aatcatatat cactaaagaa ttcacaggaa aacttttaag ttctgaagga   2340
aaacaatttg ctattactga attggaacat ccattattta atgtgataac aaacgcaacg   2400
ataaataatg tgaattttga aaatgtagag atagaacgtt ctggtcaaga taatattgca   2460
tcattagcca atactatgaa aggttcttca gttattacaa atgtcaaaat tacaggcaca   2520
cttcaggtc gtaataatgt tgctggattt gtaaataata tgaatgatgg aactcgtatt   2580
gaaaatgttg ctttctttgg caaactacac tctacaagtg gaaatggctc tcatacaggg   2640
ggaattgcag gtacaaacta tagaggaatt gttagaaaag catatgttga tgctactatt   2700
acaggaaaca aaacacgcgc cagcttgtta gttcctaaag tagattatgg attaactcta   2760
gaccatctta ttggtacaaa agctctccta actgagtcgg ttgtaaaagg taaaatagat   2820
gtttcaaatc cagtagaagt tggagcaata gcaagtaaga cttggcctgt aggtacggta   2880
agtaattctg tcagctatgc taagattatc cgtggagagg agttattcgg ctctaacgac   2940
gttgatgatt ctgattatgc tagtgctcat ataaaagatt tatatgcggt agagggatat   3000
tcgtcaggta atagatcatt taggaaatct aaaacatta ctaaattaac taaagaacaa   3060
gctgatgcta aagttactac tttcaatatt actgctgata aattagaaag tgatctatct   3120
cctcttgcaa aacttaatga agaaaaagcc tattctagta ttcaagatta taacgctgaa   3180
tataaccaag cctataaaaa tcttgaaaaa ttaataccat tctacaataa agattatatt   3240
gtatatcaag gtaataaatt aaataaagaa caccatctaa atactaaaga agttctttct   3300
gttaccgcga tgaacaacaa tgagtttatc acaaacctag atgaagctaa taaaattatt   3360
gttcactatg cggacggtac aaaagattac tttaacttgt cttctagcag tgaaggttta   3420
agtaatgtaa aagaatatac tataactgac ttaggaatta aatatacacc taatatcgtt   3480
caaaaagata acactactct tgttaatgat ataaaatcta ttttagaatc agtagagctt   3540
```

```
cagtctcaaa cgatgtatca gcatctaaat cgattaggtg actatagagt taatgcaatc   3600
aaagatttat atttagaaga aagcttcaca gatgttaaag aaaacttaac aaacctaatc   3660
acaaaattag ttcaaaacga agaacatcaa ctaaatgatt ctccagctgc tcgtcaaatg   3720
attcgtgata aagtcgagaa aaacaaagca gctttattac taggtttaac ttacctaaat   3780
cgttactatg gagttaaatt tggtgatgtt aatattaaag aattaatgct attcaaacca   3840
gatttctatg gtgaaaagt tagcgtatta gacagattaa ttgaaatcgg ttctaaagag   3900
aacaacatta aaggttcacg tacattcgac gcattcggtc aagtattggc taaatatact   3960
aaatcaggta atttagatgc attttttaaat tataatagac aattgttcac aaatatagac   4020
aatatgaacg attggtttat tgatgctaca gaagaccatg tctacatcgc agaacgcgct   4080
tctgaggtcg aagaaattaa aaattctaaa catcgtgcat tcgataattt aaaacgaagt   4140
caccttagaa atactatact cccactactg aatattgata aagcacatct ttatttaatt   4200
tcaaattata atgcaattgc ctttggtagt gcagagcgat taggtaaaaa atcattagaa   4260
gatattaaag atatcgttaa caaagctgca gatggttata gaaactatta tgatttctgg   4320
tatcgtctag cgtctgataa cgttaaacaa cgactactaa gagatgctgt tattcctatt   4380
tgggaaggtt ataacgctcc tggtggatgg gttgaaaaat atggccgcta taataccgac   4440
aaagtatata ctcctcttag agaattcttt ggtcctatgg ataagtatta taattataat   4500
ggaacaggag cttatgctgc tatatatcct aactctgatg atattagaac tgatgtaaaa   4560
tatgttcatt tagaaatggt tggtgaatac ggtatttcag tttacacaca tgaaacaaca   4620
cacgtcaacg accgtgcgat ttacttaggt ggctttggac accgtgaagg tactgatgct   4680
gaagcatatg ctcagggtat gctacaaact cctgttactg gtagtggatt tgatgagttt   4740
ggttctttag gtattaatat ggtatttaaa cgcaaaaatg atgggaatca gtggtatatt   4800
acagatccaa aaactctaaa aacacgagaa gatattaata gatatatgaa gggttataat   4860
gacactttaa ctcttcttga tgaaattgag gctgaatctg tgatttctca acaaaataaa   4920
gatttaaata gtgcatggtt caaaaaaata gatagagaat accgtgataa caataaatta   4980
aatcaatggg ataaaattcg aaatctaagt caagaagaga aaatgaatt aaatattcaa   5040
tctgttaatg atttagttga tcaacaatta atgactaatc gcaatccagg taatggtatc   5100
tataaacccg aagcaattag ctataacgat caatcacctt atgtaggtgt tagaatgatg   5160
accggtatct acggaggtaa tactagtaaa ggtgctcctg gagctgtttc attcaaacat   5220
aatgctttta gattatgggg ttactacgga tacgaaaatg ggttcttagg ttatgcttca   5280
aataaatata aacaacaatc taaaacagat ggtgagtctg ttctaagtga tgaatatatt   5340
atcaagaaaa tatctaacaa tacatttaat actattgaag aatttaaaaa agcttacttc   5400
aaagaagtta aagataaagc aacgaaagga ttaacaacat tcgaagtaaa tggttcttcc   5460
gtttcatcat acgatgattt actgacattg tttaaagaag ctgttaaaaa agatgccgaa   5520
actcttaaac aagaagcaaa cggtaataaa acagtatcta tgaataatac agttaaatta   5580
aaagaagctg tttataagaa acttcttcaa caaacaaata gctttaaaac ttcaatcttt   5640
aaa                                                               5643


SeqID 36
atgaataaac gtctattttc aaaaatgagt ctggtgacgt tgccaatttt agccttgttt     60
tcacaatcag ttttggcgga agaaaacatc catttttcga gctgtaagga agcttgggcg    120
aatggctatt cggatattca cgagggagaa cctggttatt ctgccaagtt agaccgtgat    180
catgatggtg tggcttgcga attgaaaaat gctcctaagg gtgcttttaa agcaaaacag    240
tcaacggcta ttccaaatcaa cacaagttca gcaacaacaa gtggttgggt taagcaggac    300
ggcgcttggt actactttga tggaaatgga aatctagtga aaaatgcatg gcagggaagc    360
tattacctga aagctgatgg taaaatggca cagagtgaat ggatttatga ctcttcttat    420
caagcttggt attatttgaa atcagatggt tcttatgcaa aaaatgcatg gcaaggagct    480
tattacctta aatcaaacgg taaaatggca caaggtgagt gggtttatga ttcttcttac    540
caagcatggt attacttgaa atcagatggt tcatatgctc gcaatgcatg gcaaggaaac    600
tactatttga aatcagatgg taaaatggct aaaggtgaat gggtttatga tgccacctat    660
caagcttggt attatttgac atcagatggt tcttatgctt acagtacatg gcaaggaaat    720
tactatctaa aatcagatgg taaaatggct gtcaatgaat gggttgatgg tggacgttat    780
catgttggcg ctgacggagt ttggaaggaa gttcaagcaa gtacagcttc ttctagtaat    840
jatagcaata gtgaatattc tgctgcttta ggaaaggcaa aaagttataa ttcgttattc    900
jacatgtcaa aaaacgtat gtatagacaa ttaacttctg attttgataa attttcaaat    960
jatgcagctc aatatgccat tgatcatta gatgat                             996


SeqID 37
atgaaagtaa tagatcaatt taaaaataag aaagtccttg ttttaggttt ggccaagtct     60
jgtgaatctg cagctcgttt gttggacaag ctaggtgcca ttgtgacagt aaatgatggg    120
iaacctttcg aggacaatcc agctgcccaa agtttgctgg aagaagggat caaggtcatt    180
icaggtggcc atcctttgga actcttggat gaagagtttg cccttatggt gaaaaatcca    240
jgtatcccct acaacaatcc catgattgaa aaggctttgg ccaaggtaat tccagtcttg    300
ictgaggtgg aattggctta tttgatttca gaagcaccga ttattggtat cacaggatcg    360
.acggtaaga caaccacaac gactatgatt ggggaagttt tgactgctgc tggccaacat    420
jtctttat cagggaatat cggctatcca gctagtcagg ttgctcaaat agcatcagat    480
.aggacacgc ttgttatgga actttcttct ttccaactca tgggtgttca agaattccat    540
cagagattg cggttattac caacctcatg ccaactcata tcgactacca tgggtcattt    600
cggaatatg tagcagccaa gtggaatatc cagaacaaga tgacagcagc tgatttcctt    660
```

```
gtcttgaact ttaatcaaga cttggcaaaa gacttgactt ccaagacaga agccactgtt    720
gtaccatttt caacacttga aaaggttgat ggagcttatc tggaagatgg tcaactctac    780
ttccgtggtg aagtagtcat ggcagcgaat gaaatcggtg ttccaggtag ccacaatgtg    840
gaaaatgccc ttgcgactat tgctgtagcc aagcttcgtg atgtggacaa tcaaaccatc    900
aaggaaactc tttcagcctt cggtggtgtc aaacaccgtc tccagtttgt ggatgacatc    960
aaggtgtta aattctataa cgacagtaaa tcaactaata tcttggctac tcaaaaagcc   1020
ttgtcaggat ttgacaacag caaggtcgtc ttgattgcag gtggtttgga ccgtggcaat   1080
gagtttgacg aattggtgcc agacattact ggactcaaga agatggtcat cctgggtcaa   1140
tctgcagaac gtgtcaaacg ggcagcagac aaggctggtg tcgcttatgt ggaggcgaca   1200
gatattgcag atgcgacccg caaggcctat gagcttgcga ctcaaggaga tgtggttctt   1260
cttagtcctg ccaatgctag ctgggatatg tatgctaact ttgaagtacg tggcgacctc   1320
tttatcgaca cagtagcgga gttaaaagaa                                   1350

SeqID 38
atgaagaaaa aatttgccct atcgtttgtg gcgcttgcaa gtgtagcact tcttgcagcc     60
tgtggagaag tgaagtctgg agcagtcaac actgctggta actcagtaga ggaaaagaca    120
attaaaatcg ggtttaactt tgaagaatca ggttctttag ctgcatacgg aacagctgaa    180
caaaaaggtg cccaattggc tgttgatgaa atcaatgccg caggtggtat cgatggaaaa    240
caaatcgaag tagtcgataa agataataag tctgaaacag ctgaggctgc ttcagttaca    300
actaaccttg taacccaatc taaagtatca gcagtcgtag gacctgcgac atctggtgcg    360
actgcagctg cggtagcgaa cgctacaaaa gcaggtgttc cattgatctc accaagtgcg    420
actcaagatg gattgactaa aggtcaagat tacctcttta ttggaacttt ccaagatagc    480
ttccaaggaa aaattatctc aaactatgtt tctgaaaaat aaatgctaa gaaagttgtt    540
ctttacactg acaatgccag tgactatgct aaagggattg caaaatcttt ccgcgagtca    600
tacaagggtg aaatcgttgc agatgaaact ttcgtagcag gtgacacaga cttccaagca    660
gcccttacaa aaatgaaagg gaaagacttt gatgctatcg ttgttcctgg ttactataat    720
gaggctggta aaattgtaaa ccaagcgcgt ggcatgggaa ttgacaaacc aatcgttggt    780
ggtgatggat caacggtga ggagtttgta caacaagcaa ctgctgaaaa agcatcaaac    840
atctacttta tctcaggctt ctcaactact gtagaagttt cagctaaagc taaagccttc    900
cttgacgctt accgtgctaa gtacaatgaa gagccttcaa catttgcagc cttggcttat    960
gattcagttc accttgtagc aaacgcagca aaaggtgcta aaaattcagg tgaaatcaag   1020
aataaccttg ctaaaacaaa agattttgaa ggtgtaactg gtcaaacaag cttcgatgca   1080
gaccacaaca cagtcaaaac tgcttacatg atgaccatga acaatggtaa agttgaagca   1140
gcagaagttg taaaaacca                                               1158

SeqID 39
atgagtattt tagaagttaa aaatctgagt cacggttttg gtgaccgtgc aattttttgaa     60
gatgtgtcct tccgtctcct caagggagaa catatcggcc tggtcggtgc caatggtgaa    120
ggaaaatcaa cctttatgag tatcgtgact ggtaaaatgc tgccagatga aggaaaggtt    180
gagtggtcca aatatgtgac ggctggttac ttggatcagc actctgtcct tgctgaaaga    240
cagtcggtgc gtgatgttct ccgtacggct tttgatgagc ttttcaaagc tgaagctcgt    300
atcaatgacc tctatatgaa aatggctgaa gacggcgcgg atgttgatgc tctcatggaa    360
gaagtaggag aacttcaaga ccgtctggag agtcgtgatt tctatacctt ggatgctaag    420
attgacgaag tagcgcgtgc tcttggtgtt atggactttg gcatggatac ggatgtaact    480
tctttgtcag gtgggcaaag aaccaaggtg cttttggcaa aacttctcct tgaaaagcct    540
gatatcttgc tgttggacga gccgaccaac tacttggatg ctgagcatat tgattggctc    600
aagcgctatc tccaaaacta tgagaatgcc tttgttctca tttcgcacga tattccattc    660
ctcaatgacg ttattaatat tgtctatcat gtggaaaatc aacagctgac gcgttactct    720
ggtgactact accagttcca agaagtttat gctatgaaga aatctcagct agaggcagcc    780
tacgaacgcc agcagaaaga gattgcagac ctcaaggact ttgtggctcg taataaagcc    840
cgtgttgcaa ctcgtaatat ggctatgtct cgtcaaaaga aattggataa gatggatatt    900
atcgaactcc aaagtgagaa accaaaacca tcctttgatt tcaaaccagc tcgtacacca    960
gggcgcttta tcttccaagc caagaacttg caaattggtt acgaccgtcc tcttactaag   1020
cctttaaatc ttaccttcga acgcaatcaa aaggttgcga ttattggtgc taatggtatt   1080
ggaaaaacaa ctctcttgaa gagtctcttg ggcattatct cgccaatcgc tggggaagtg   1140
gagcgtggag attatttaga acttggttat tttgagcagg aagtagaagg cggtaatcgc   1200
caaactcctc ttgaagctgt ctggaatgcc tttcctgccc ttaatcaagc agaagtccgt   1260
tcagcccttg cccgttgtgg tttgacaacc aaacatattg aaagccagat tcaagtatta   1320
tcaggggag agcaagccaa ggttcgtttc tgtctcttga tgaatcgtga aaacaacgtt   1380
tagtgctgg acgagccgac caaccatttg gatgtggatg caaaggatga gctcaaacgc   1440
tctctcaaag aatatagggg atctatcctt atggtctgcc acgagccaga cttttatgaa   1500
ggctggatag accaaatatg ggatttaat aatttaact                          1539

SeqID 40
atgaagaaaa agaatggtaa agctaaaaag tggcaactgt atgcagcaat cggtgctgcg     60
gtgtagttg tattgggtgc tgggggatt ttactcttta gacaaccttc tcagactgct    120
taaaagatg agcctactca tcttgttgtt gccaaggaag gaagcgtggc ctcctctgtt    180
tattgtcag ggacagtaac agcaaaaaat gaacaatatg tttattttga tgctagtaag    240
```

```
ggtgatttag atgaaatcct tgtttctgtg ggcgataagg tcagcgaagg gcaggcttta    300
gtcaagtaca gtagttcaga agcgcaggcg gcctatgatt cagctagtcg agcagtagct    360
agggcagatc gtcatatcaa tgaactcaat caagcacgaa atgaagccgc ttcagctccg    420
gctccacagt taccagcgcc agtaggagga gaagatgcaa cggtgcaaag cccaactcca    480
gtggctggaa attctgttgc ttctattgac gctcaattgg gtgatgcccg tgatgcgcgt    540
gcagatgctg cggcgcaatt aagcaaggct caaagtcaat tggatgcaac aactgttctc    600
agtaccctag agggaactgt ggtcgaagtc aatagcaatg tttctaaatc tccaacaggg    660
gcgagtcaag ttatggttca tattgtcagc aatgaaaatt tacaagtcaa gggagaattg    720
tctgagtaca atctagccaa cctttctgta ggtcaagaag taagctttac ttctaaagtg    780
tatcctgata aaaaatggac tgggaaatta agctatattt ctgactatcc taaaaacaat    840
ggtgaagcag ctagtccagc agccgggaat aatacaggtt ctaaataccc ttatactatt    900
gatgtgacag gcgaggttgg tgatttgaaa caaggttttt ctgtcaacat tgaggttaaa    960
agcaaaacta aggctattct tgttcctgtt agcagtctag taatggatga tagtaaaaat   1020
tatgtctgga ttgtggatga acaacaaaag gctaaaaaag ttgaggtttc attgggaaat   1080
gctgacgcag aaaatcaaga aatcacttct ggtttaacga acggtgctaa ggtcatcagt   1140
aatccaacat cttccttgga agaaggaaaa gaggtgaagg ctgatgaagc aactaat      1197
```

SeqID 41
```
tcagaaacaa atcacgaaat tgattcaaat tttgcaggtc gtttaaatat cctgcgtgcg     60
ggtgttcttg atgctaacga tggaattatt tccattgctg gtgtggttat cggagttgcc    120
agtgccacga ccaatatctg gattatcttt ttatcaggtt ttacggctat cttagctggt    180
gccttttcaa tggctggtgg agaatatgta tccgtttcaa ctccaaaaga taccgaggaa    240
gctgccgttt cgcgagaaaa actcttgcta gaccaagata gggaactagc caaaaaatcc    300
ctctatgctg cttatatcca aaatggagaa ttcaaaactt ctgcccaact cttgaccaat    360
aagatctttc ttaaaaatcc actcaaggct gagtagagga aaaatatggt gattgagtat    420
gaagaattta ccaatccttg gcacgctgcc atttctagct tcgttgcctt tttccttaga    480
agtttgcctc caatgctgtc agtgaccatt tttccaagtg attaccgcat ccctgctacc    540
gtccttattg tcggtgtggc ccttcttctc actggttaca caagtgctag acttggaaaa    600
gccccaacca aaacagctat gattcggaac cttgctattg gtctcttgac catgggagtt    660
accttcctgc tcggacaact tttcagcatt                                     690
```

SeqID 42
```
atgaaaaaga aattaactag tttagcactt gtaggcgctt tttaggtttt gtcatggtat     60
gggaatgttc aggctcaaga aagttcagga aataaaatcc actttatcaa tgttcaagaa    120
ggtggcagtg atgcgattat tcttgaaagc aatggacatt ttgccatggt ggatacagga    180
gaagattatg atttcccaga tggaagtgat tctcgctatc catggagaga aggaattgaa    240
acgtcttata agcatgttct aacagaccgt gtctttcgtc gtttgaagga attgggtgtc    300
caaaaacttg attttatttt ggtgacccat acccacagtg atcatattgg aaatgttgat    360
gaattactgt ctacctatcc agttgaccga gtctatctta agaaatatag tgatagtcgt    420
attactaatt ctgaacgtct atgggataat ctgtatggct atgataaggt tttacagact    480
gctgcagaaa aaggtgtttc agttattcaa aatatcacac aaggggatgc tcattttcag    540
tttggggaca tggatattca gctctataat tatgaaaatg aaactgattc atcgggtgaa    600
ttaaagaaaa tttgggatga caattccaat tccttgatta gcgtggtgaa agtcaatggc    660
aagaaaattt accttggggg cgatttagat aatgttcatg gagcagaaga caagtatggt    720
cctctcattg gaaaagttga tttgatgaag tttaatcatc accatgatac caacaaatca    780
aataccaagg atttcattaa aaatttgagt ccgagtttga ttgttcaaac ttcggatagt    840
ctaccttgga aaaatggtgt tgatagtgag tatgttaatt ggctcaaaga acgaggaatt    900
gagagaatca acgcagccag caaagactat gatgcaacag ttttgatat tcgaaaagac    960
ggttttgtca atatttcaac atcctacaag ccgattccaa gttttcaagc tggttggcat   1020
aagagtgcat atgggaactg gtggtatcaa gcgcctgatt ctacaggaga gtatgctgtc   1080
ggttggaatg aaatcgaagg tgaatggtat tactttaacc aaacgggtat cttgttacag   1140
aatcaatgga aaaaatggaa caatcattgg cagactctgg ttctatttgct   1200
aaaaattgga agaaaatcgc tggaatctgg tattattta acaaagaaaa ccagatggaa   1260
attggttgga ttcaagataa agagcagtgg tattatttgg atgttgatgg ttctatgaag   1320
acaggatggc ttcaatatat ggggcaatgg tattactttg ctccatcagg ggaaatgaaa   1380
atgggctggg taaaagataa agaaacctgg tactatatgg attctactgg tgtcatgaag   1440
acaggtgaga tagaagttgc tggtcaacat tattatctgg aagattcagg agctatgaag   1500
caaggctggc ataaaaaggc aaatgattgg tatttctaca agacagacgg ttcacgagct   1560
gtgggttgga tcaaggacaa ggataaatgg tacttcttga agaaaatgg tcaattactt   1620
gtgaacggta agacaccaga aggttatact gtggattcaa gtggtgcctg gttagtggat   1680
gtttcgatcg agaaatctgc tacaattaaa actacaagtc attcagaaat aaaagaatcc   1740
aaagaagtag tgaaaaagga tcttgaaaat aaagaaacga gtcaacatga aagtgttaca   1800
aatttttcaa ctagtcaaga tttgacatcc tcaacttcac aaagctctga aacgagtgta   1860
aacaaatcgg aatcagaaca g                                             1881
```

SeqID 43
```
atggacttag gtcccacca aagaggtatt agtgtcgtgt ctcaatctta tatcaatgtt     60
atcggtgctg gtttggcagg ttctgaagca gcttaccaaa tcgcagagcg tggtattcca   120
```

```
gttaaactat atgaaatgcg tggtgtcaag tctacacccc agcataaaac agacaatttt      180
gctgagttgg tttgttccaa ttctttgcgt ggggatgctt tgacaaatgc agttggtctt      240
ctcaaggaag aaatgcgtcg cttgggttct gttatcttgg aatctgctga ggctacacgt      300
gttcctgcag gtggtgccct tgcagtggac cgtgatggtt tctctcaaat ggtgaccgaa      360
aaagttgcca accacccctt gattgaagtg gttcgtgatg aaattacaga attgccgaca      420
gatgttatta cggttatcgc tactggtcct ttgacaagtg atgccttggc tgaaaagatt      480
catgctctta atgacggtgc tggttttat ttctacgatg cggcagcgcc tattatcgat      540
gtcaacacta tcgatatgag caaggtctac ctcaaatcac gttatgataa gggagaagcg      600
gcctacctca atgcccctat gaccaagcaa gaatttatgg atttccatga agctttggtc      660
aatgcagaag aagcaccgct tagttctttt gaaaaagaaa agtactttga aggatgtatg      720
cctatcgaag tcatggccaa acgtggcatt aaaactatgc tttatgtccc tatgaagcca      780
gtcggtcttg agtacccaga cgactataca ggacctcgtg atggagaatt taaaacacct      840
tatgcggttg tgcaacttcg tcaggataat gcagctggta gcctctacaa tattgttggt      900
ttccagaccc acctcaaatg gggagaacaa aagcgtgtct tccaaatgat tccgggtctt      960
gaaaatgcgg agtttgtccg ttatggtgtg atgcatcgca attcttacat ggattcacca     1020
aatcttcttg agcagactta ccgttctaag aaacaaccaa atctcttctt tgctggtcaa     1080
atgacgggtg tggaaggcta tgttgagtcg gcggcttcag gcttagttgc gggaattaac     1140
gcagctgtc tcttcaagga agaaagcgag gctattttcc ccgagacgac agcgattgga     1200
agcttagctc attacattac ccatgccgac agcaaacatt tccaaccaat gaatgtcaat     1260
tttgggatca tcaaggagtt ggaaggcgag cgtatccgtg ataagaaggc tcgttatgaa     1320
aaaattgcag agcgtgccct tgccgactta gaggaatttt tgactgtc                  1368
```

SeqID 44
```
atgttaatcg gaatcccaaa agaaattaaa aataacgaaa accgtgtcgc cctcacacct      60
gcaggtgttc atagcttagt tagtcgtggt catcgtgtcc ttatcgaaac aaatgctggt      120
ctcggttctg gctttactga tgctgactat caaaagcaag gagctgagat tgtcgctact      180
gctggtgaag cttgggcagc agagttggtt gtgaaagtaa aagaatcttt aagttctgaa      240
tacggttact tgcgcgacga tcttcttctc ttcacctact tgcacatggc cgctgctcca      300
gaattagcag atgctatgtt aacagcaaaa acaactgaaa ctgttcgtga caatcaagga      360
caactaccgc tcctcgttcc tatgagtgag gttgcaggtc gtatggctgt tcaaatcgga      420
gctcacttcc ttactaagca agctggtggc tctggtgttc tacttggtgg tgtaccaggt      480
gttccaaaag gaaaagtaac tatcatcggt ggtggtgtcg tcggtacaca tgctgcccgc      540
atcgcccttg gtcttggtgc tcaagtgact attttagata ttagttccaa gcgtctctca      600
gttctagaag aagtctttgg aagtcaaatt caaactctta tgtctaattc attcaacatt      660
gaagcaagtg tgagagatgc tgatgtggtg attggagcca ttctcatccc tggtgcaaaa      720
gcaccggaat tggtgacaga tgagatgtc aaacaaatgc gtccaggctc tgtatcgttg      780
acgttgctgt tgaccaaggt ggcgttatcg aaacagctga ccgtgtgaca acgcacgatg      840
aacccgtcta tgaaaaacac ggtgttctcc actatgccgt tgccaatatc cctggtgcgg      900
ttgctcgcac ttcaaccatc gccctaacca atgtcactct tccttatatc gaagctttgg      960
ctggcaaagg attcgcacaa gcaatctctg aagatgaagg cttgcgtcaa ggtgtgacta     1020
cttatcaagg ttacttgact aacctaccag ttgctcaagg acttaatcgt gactacactg     1080
atatcaatga tttagta                                                     1097
```

SeqID 45
```
atgaaaatta ataaaaaata tctagcaggt tcagtggcag tccttgccct aagtgtttgt      60
tcctatgaac ttggtcgtca ccaagctggt caggttaaga aagagtctaa tcgagtttct      120
tatatagatg gtgatcaggc tggtcaaaag gcagaaaact tgacaccaga tgaagtcagt      180
aagagggagg ggatcaacgc cgaacaaatc gtcatcaaga ttacggatca aggttatgtg      240
acctctcatg gagaccatta tcattactat aatggcaagg tcccttatga tgccatcatc      300
agtgaagagc tcctcatgaa agatccgaat tatcagttga aggattcaga cattgtcaat      360
gaaatcaagg gtggttatgt tatcaaggta gatggaaaat actatgttta ccttaaggat      420
gcagctcatg cggataatat tcggacaaaa gaagagatta aacgtcagaa gcaggaacac      480
agtcataatc acggggtgg ttctaacgat caagcagtag ttgcagccag agcccaagga      540
cgctatacaa cggatgatgg ttatatcttc aatgcatctg atatcattga gacacgggtg      600
gatgcttata tcgttcctca cggcgaccat taccattaca ttcctaagaa tgagttatca      660
gctagcgagt tagctgctgc agaagcctat tggaatggga agcagggatc tcgtccttct      720
tcaagttcta gttataatgc aaatccagct caaccaagat tgtcagagaa ccacaatctg      780
actgtcactc caacttatca tcaaaatcaa ggggaaaaca tttcaagcct tttacgtgaa      840
ttgtatgcta aacccttatc agaacgccat gtggaatctg atggccttat tttcgaccca      900
gcgcaaatca caagtcgaac cgccagaggt gtagctgtcc ctcatggtaa ccattaccac      960
tttatccctt atgaacaaat gtctgaattg gaaaaacgaa ttgctcgtat tattcccctt     1020
cgttatcgtt caaaccgttg ggtaccagat tcaagaccaa aacaaccaag tccacaatcg     1080
actccggaac ctagtccaag tccgcaacct gcaccaaaatc ctcaaccagc tccaagcaat     1140
ccaattgatg agaaattggt caaagaagct gttcgaaaag taggcgatgg ttatgtcttt     1200
gaggagaatg gagtttctcg ttatatccca gccaaggatc tttcagcaga aacagcagca     1260
ggcattgata gcaaactggc caagcaggaa agtttatctc ataagctagg agctaagaaa     1320
actgacctcc catctagtga tcgagaattt acaataaggt cttatgactt actagcaaga     1380
attcaccaag atttacttga taataaaggt cgacaagttg attttgaggc tttggataac     1440
```

```
ctgttggaac gactcaagga tgtcccaagt gataaagtca agttagtgga tgatattctt   1500
gccttcttag ctccgattcg tcatccagaa cgtttaggaa aaccaaatgc gcaaattacc   1560
tacactgatg atgagattca agtagccaag ttggcaggca agtacacaac agaagacggt   1620
tatatctttg atcctcgtga tataaccagt gatgagggggg atgcctatgt aactccacat   1680
atgacccata gccactggat taaaaaagat agtttgtctg aagctgagag agcggcagcc   1740
caggcttatg ctaaagagaa aggtttgacc cctccttcga cagaccatca ggattcagga   1800
aatactgagg caaaaggagc agaagctatc tacaaccgcg tgaaagcagc taagaaggtg   1860
ccacttgatc gtatgcctta caatcttcaa tatactgtag aagtcaaaaa cggtagttta   1920
atcatacctc attatgacca ttaccataac atcaaatttg agtggtttga cgaaggcctt   1980
tatgaggcac ctaaggggta tactcttgag gatcttttgg cgactgtcaa gtactatgtc   2040
gaacatccaa acgaacgtcc gcattcagat aatggttttg gtaacgctag cgaccatgtt   2100
cgtaaaaata aggtagacca agacagtaaa cctgatgaag ataaggaaca tgatgaagta   2160
agtgagccaa ctcaccctga atctgatgaa aaagagaatc acgctggttt aaatccttca   2220
gcagataatc tttataaacc aagcactgat acggaagaga cagaggaaga agctgaagat   2280
accacagatg aggctgaaat tcctcaagta gagaattctg ttattaacgc taagatagca   2340
gatgcggaag ccttgctaga aaaagtaaca gatcctagta ttagacaaaa tgctatggag   2400
acattgactg gtctaaaaag tagtcttctt ctcggaacga aagataataa cactatttca   2460
gcagaagtag atagtctctt ggctttgtta aaagaaagtc aaccggctcc tatacag      2517
```

SeqID 46

```
atgaaattta gtaaaaaata tatagcagct ggatcagctg ttatcgtatc cttgagtcta     60
tgtgcctatg cactaaacca gcatcgttcg caggaaaata aggacaataa tcgtgtctct    120
tatgtggatg gcagccagtc aagtcagaaa agtgaaaact tgacaccaga ccaggttagc    180
cagaaagaag gaattcaggc tgagcaaatt gtaatcaaaa ttacagatca gggctatgta    240
acgtcacacg gtgaccacta tcattactag aatgggaaag ttccttatga tgccctcttt    300
agtgaagaac tcttgatgaa ggatccaaac tatcaactta aagacgctga tattgtcaat    360
gaagtcaagg gtggttatat catcaaggtc gatggaaaat attatgtcta cctgaaagat    420
gcagctcatg ctgataatgt tcgaactaaa gatgaaatca atcgtcaaaa acaagaacat    480
gtcaaagata atgagaaggt taactctaat gttgctgtag caaggtctca gggacgatat    540
acgacaaatg atggttatgt ctttaatcca gctgatatta tcgaagatac gggtaatgct    600
tatatcgttc ctcatggagg tcactatcac tacattccca aaagcgattt atctgctagt    660
gaattagcag cagctaaagc acatctggct ggaaaaaata tgcaaccgag tcagttaagc    720
tattcttcaa cagctagtga caataacacg caatctgtag caaaaggatc aactagcaag    780
ccagcaaata aatctccagat tctttgaagg aactctatga ttcacctagc    840
gcccaacgtt acagtgaatc agatggcctg gtctttgacc ctgctaagat tatcagtcgt    900
acaccaaatg gagttgcgat tccgcatggc gaccattacc actttattcc ttacagcaag    960
ctttctgcct tagaagaaaa gattgccaga atggtgccta tcagtggaac tggttctaca   1020
gtttctacaa atgcaaaacc taatgaagta gtgtctagtc taggcagtct ttcaagcaat   1080
ccttcttctt taacgacaag taaggagctc tcttcagcat ctgatggtta tatttttaat   1140
ccaaaagata tcgttgaaga aacggctaca gcttatattg taagacatgg tgatcatttc   1200
cattacattc caaaatcaaa tcaaattggg caaccgactc ttccaaacaa tagtctagca   1260
acaccttctc catctcctcc aatcaatcca ggaacttcac atgagaaaca tgaagaagat   1320
ggatacggat ttgatgctaa tcgtattatc gctgaagatg aatcaggttt tgtcatggag   1380
cacggagacc acaatcatta tttcttcaag aaggacttga cagaagagca aattaaggct   1440
gcgcaaaaac atttagagga agttaaaact agtcataatg gattagattc tttgtcatct   1500
catgaacagg attatccaag taatgccaaa gaaatgaaag atttagataa aaaaatcgaa   1560
gaaaaaattg ctggcattat gaaacaatat ggtgtcaaac gtgaaagtat tgtcgtgaat   1620
aaagaaaaaa atgcgattat ttatccgcat ggagatcacc atcatgcaga tccgattgat   1680
gaacataaac cggttggaat tggtcattct cacagtaact atgaactgtt taaacccgaa   1740
gaaggagttg ctaaaaaaga agggaataaa gtttatactg gagaagaatt aacgaatgtt   1800
gttaatttgt taaaaaatag tacgtttaat aatcaaaaact ttactctagc caatggtcaa   1860
aaacgcgttt cttttagttt tccgcctgaa ttggagaaaa aattaggtat caatatgcta   1920
gtaaaattaa taacaccaga tggaaaagta ttggagaaag tatctggtaa agtatttgga   1980
gaaggagtag ggaatattgc aaactttgaa ttagatcaac cttatttacc aggacaaaca   2040
tttaagtata ctatcgcttc aaaagattat ccagaagtaa gttatgatgg tacatttaca   2100
gttccaacct cttttagctta caaaatggcc agtcaaacga ttttctatcc tttccatgca   2160
gggggatactt atttaagagt gaaccctcaa tttgcagtgc ctaaaggaac tgatgcttta   2220
gtcagagtgt ttgatgaatt tcatggaaat gcttatttag aaaataacta taaagttggt   2280
gaaatcaaat taccgattcc gaaattaaac caaggaacaa ccagaacggc cggaaataaa   2340
attcctgtaa cctctcatgg caaatgcttat ttggacaatc aatcgactta tattgtggaa   2400
gtaacctatct tggaaaaaga aaatcaaact gataaaccaa gtattctacc acaatttaaa   2460
aggaataaag cacaagaaaa cttaaaaactt gatgaaaagg tagaagaacc aaaagactagt   2520
gagaaggtag aaaaagaaaa actttctgaa actgggaata gtactagtaa ttcaacgtta   2580
gaagaagttc ctacagtgga tcctgtacaa gaaaaagtag caaaatttgc tgaaagttat   2640
gggatgaagc tagaaaatgt cttgtttaat atggacggaa caattgaatt atatttacca   2700
tcggagaagt cattaaaaaa gaatatggca gattttacag gagaagcacc tcaaggaaat   2760
ggtgaaaata aaccatctga aaatgaaaaa gtatctactg gaacagttga gaaccaacca   2820
acagaaaata aaccagcaga ttctttacca gaggcaccaa acgaaaaacc tgtaaaacca   2880
```

```
gaaaactcaa cggataatgg aatgttgaat ccagaaggga atgtggggag tgaccctatg    2940
ttagatccag cattagagga agctccagca gtagatcctg tacaagaaaa attagaaaaa    3000
tttacagcta gttacggatt aggcttagat agtgttatat tcaatatgga tggaacgatt    3060
gaattaagat tgccaagtgg agaagtgata aaaaagaatt tatctgatct catagcg      3117
```

SeqID 47
```
atgaaaattt tatttgtagc agcagagggt gcaccctttt caaaaacagg tggtttggga     60
gacgtcattg gcgctcttcc aaaatcactg gtaaaagctg ggcacgaagt tgcagtgatt    120
ttaccctact atgatatggt agaggctaaa tttggaaatc agattgaaga tgtgcttcat    180
tttgaggtga gcgttggttg gcgcagacag tattgtggaa ttaagaaaac agtattaaat    240
ggtgtaacct tctactttat tgacaatcaa tattatttct tccgtggtca tgtttacggt    300
gattttgatg acggagaacg ctttgccttt ttccaactgg ctgccattga ggctatggaa    360
aggattgact ttattcctga tcttctccat gttcatgact accatacagc tatgattcct    420
ttcttgttga aggaaaaata ccgttggatt caagcctatg aggacattga aacagtttta    480
accattcata atttagaatt ccaaggacaa tttttcagaag gaatgttggg tgatttgttt    540
ggagttggct ttgaacgtta cgctgatggc acccttcgat ggaacaactg tctgaactgg    600
atgaaggcag gtattctcta tgcgaaccgt gtttcaaccg tttcacctag ctatgctcat    660
gaaattatga ctagtcagtt tggatgtaat ttggatcaga ttcttaaaat ggagtctggt    720
aaagtatctg gtatcgtgaa tgggattgat gctgatcttt ataatcctca gacggatgct    780
ctttttagact atcatttcaa tcaggaagat ttgtctggga aagccaaaaa taaggcaaaa    840
ttgcaagaaa gagttggctt gcctgttaga gcagacgttc cactggtggg aattgtttct    900
cgtttgacac gtcaaaaagg ttttgatgtg gtggtcgaaa gtcttcacca tatcttgcaa    960
gaagatgttc agattgttct tttgggaact ggcgatccag cctttgaagg agctttctca   1020
tggtttgctc agatttaccc agacaagcta tcaacaaata tcactttga tgtcaaactt   1080
gctcaggaaa tctacgctgc ttgtgacctc ttcctcatgc caagtcgttt tgaaccgtgt   1140
ggcttgtctc aaatgatggc tatgcgttat ggaaccttgc cattggtcca tgaagttgga   1200
ggcttgcgag atacgggtac cgcttcaat ccaatcgaag gagcggtac tggctttagc    1260
tttgacaatc tatctcctta ttggttaaat tggactttcc aaacagcatt ggacttgtat   1320
agaaaccatc cagacatttg gagaaaccta caaaaacaag ctatggagag tgacttctca   1380
tgggatacag cctgcaagtc ataccttgac ttgtaccata gtttagttaa t          1431
```

SeqID 48
```
atggaaaagt attttggtga aaaacaagag cgttttttcat ttagaaaatt atcagtagga     60
cttgtatctg caacgatttc aagtttattt tttatgtctg tattagctag ttcatctgtg    120
gatgctcaag aaactgcggg agttcactat aaatatgtgg cagattcaga gctatcatca    180
gaagaaaaga agcagcttgt ctatgatatt ccgagaatga tgatgaaact    240
tattatcttg tttataagtt aaattctcaa aatcaactga cggaattgcc aaatactgga    300
agcaagaatg agaggcaagc cctagttgct ggtgctagct tagctgctat gggaattttta    360
attttttgctg tttccaagaa aaaggttaag aataaaacgg tattacattt agtattggtt    420
gcagggatag aaatggtgt cttagtttca gtccatgctt tagaaaatca tcttttgcta    480
aattacaata cggactatga attgacctct ggagaaaaat tacctcttcc taaagagatt    540
tcaggttaca cttatattgg atatatcaaa gaggaaaaaa cgacttctga gtctgaagta    600
agtaatcaaa agagttcagt tgccactcct acaaaacaac aaaaggtgga ttataatgtt    660
acaccgaatt ttgtagacca tccatcaaca gtacaagcta ttcaggaaca aacacctgtt    720
tcttcaacta agcagcctgt agttcaagta gttgaaaaac ctttctac tgaattaatc    780
aatccaagaa aagaagagaa acaatcttca gattctcaag aacaattagc cgaacataag    840
aatctagaaa cgaagaaaga ggagaagatt tctccaaaag aaaagactgg ggtaaataca    900
ttaaatccac aggatgaagt tttatcaggt caattgaaca aacctgaact cttatatcgt    960
gaggaaacta tggagacaaa aatagatttt caagaagaaa ttcaagaaaa tcctgattta   1020
gctgaaggaa ctgtaagagt aaaaacaagaa ggtaaattag gtaagaaagt tgaaatcgtc   1080
agaatattct ctgtaaacaa ggaagaagtt tcgcgagaaa ttgtttcaac ttcaacgact   1140
gcgcctagtc caagaatagt cgaaaaaggt actaaaaaaa ctcaagttat aaaggaacaa   1200
cctgagactg gtgtagaaca taaggacgta cagtctggag ctattgttga accegcaatt   1260
cagcctgagt tgcccgaagc tgtagtaagt gacaaaggcg aaccagaagt tcaacctaca   1320
ttacccgaag cagttgtgac cgacaaaggt gagactgagg ttcaaccaga gtcgccagat   1380
actgtggtaa gtgataaagg tgaaccagag caggtagcac cgcttccaga atataagggt   1440
aatattgagc aagtaaaacc tgaaactccg gttgagaaga ccaaagaaca aggtccagaa   1500
aaaactgaag aagttccagt aaaaccaaca gaagaaacac cagtaaatcc aaatgaaggt   1560
actacagaag aacctcaat tcaagaagca gaaaatccag ttcaacctgc agaagaatca   1620
acaacgaatt cagagaaagt atcaccagat acatctagca aaaatactgg ggaagtgtcc   1680
agtaatccta gtgattcgac aacctcagtt ggagaatcaa ataaaccaga acataatgac   1740
tctaaaaatg aaaattcaga aaaaactgta gaagaagttc cagtaaatcc aaatgaaggc   1800
acagtagaag gtacctcaaa tcaagaaaca gaaaaaccag ttcaacctgc agaagaaaca   1860
caaacaaact ctgggaaaat agctaacgaa aatactggag aagtatccaa taaacctagt   1920
gattcaaaac caccagttga agaatcaaat caaccagaaa aaacggaac tgcaacaaaa   1980
ccagaaaatt caggtaatac aacatcagag aatggacaaa cagaaccaga accatcaaac   2040
ggaaattcaa ctgaggatgt ttcaaccgaa tcaaacacat ccaattcaaa tggaaacgaa   2100
gaaattaaac aagaaaatga actagaccct gataaaaagg tagaagaacc agagaaaaca   2160
```

```
cttgaattaa gaaatgtttc cgacctagag ttatacagtt tgtcaaatgg tacttataaa    2220
caacacattt cgttagagca agttccaagc aatccaaata gctactttgt taaagtgaaa    2280
tcttcttcat tcaaagatgt atacctacca gtagcatcaa tatcagagga aagaaaaaat    2340
gataaaatcc tttataaaat cacagcaaaa gtagagaagc ttcagcagga gatagaaagc    2400
agatataaag ataattttac cttctatcta gctaagaagg gaacagaaga aacaacaaac    2460
tttacttcct ttagtaatct ggtcaaagct ataaaccaaa atccctctgg aacctatcat    2520
ttagcggcca gcctgaatgc taacgaagtg gagcttggtc ctgatgaaag atcctatatc    2580
aaggacacct ttactggtcg tttaatcggt gaaaaagatg gcaagaatta tgctatctat    2640
aatttgaaaa aacctctgtt tgaaaacttg agtggtgcta cagtagaaaa actgagtcta    2700
aaaaatgttg ctatttcagg gaaagatgat atcggttcac tggcaaatga agctcagaat    2760
aacacaaaaa ttaagcaagt tcacgtcgat ggtgttctgg ctggtgaacg tggtatcggt    2820
ggtttgctgg ctaaggctga gcaatcaagc atcacagaga gcagtttcaa gggaagaatt    2880
atcaacactt atgaaacgac tgctgcctac aatatcggtg gtatggtcgg tcatttgaca    2940
ggtgacaagg ctttacttac taagtcaaaa gcgacagtag ccatttcatc taacacaaat    3000
acttcagatc agactgtggg tggacttgca ggcctagtag accgagatgc acagatccaa    3060
gatagctatc ctgaaggtga tatcaacaat gtcaagcact ttggtagagt cgctggagtg    3120
gcaggcaatt tgtgggatcg aacttctggt gatgttaggc atgctggaag tttgaccaat    3180
gttctcagcg atgttaatgt aaccaacgga aatgccatca ctggttacca ctataacgaa    3240
atgaaggtaa aggacacatt cagcagcaag gccaacagag tctacaatgt caccttggtc    3300
aaggatgagg tcgtcagcaa ggaatccttt gaagaaagag gaacaatgct agatgcttct    3360
caaattgcaa gcaaaaaagc agaaatcaat cctctcattt taccaacagt ggagccactt    3420
tcaacaagtg gcaaaaaaga cagtgatttt tctaaggtgg cctattatca agctaagcgc    3480
aacttgactt ataaaaacat tgaaaaattg ctacctttct acaacaaggc aaccatcgtc    3540
aaatacggaa acctggtcaa tgagaacagt cttttatatc aaaaagaact cttgtcagca    3600
gtcatgatga aggacaacca agtcatcaca gacattgttt ctaacaaaac gactgcaaac    3660
aaactcttgc ttcactacaa ggatgattta tctgagaagc tggatctcaa ataccagaat    3720
gatttcgcca aattagcaga atatagtctg ggcaatactg gacttctcta tacgccaaac    3780
caattcctgt atgaccaaac ctctatcatc aagcaagtct tacctgactt acaaaaaggtt    3840
gactatcatt cagaagccat cagaaagacg ctgggtattt ctccaaacgt caagcaaact    3900
gagctctatc tagaagacca gttcgccaaa acaaaacaac aactggaaga cagtttgaaa    3960
aaactcttgt cagcggatgc tggactggct agtgctaacc ccgtcactga aggttatctt    4020
gtagataaaa tcaaacgcaa caaggaagcc ttgctacttg gcttgaccta tctggaacgg    4080
tggtataact ttagctatgg tcaggtgaat gtcaaagacc tagttctgta ccatttggac    4140
ttctttggta aggggaatgc ttcaccatta gatactctga ttgagttgga taaatctggc    4200
tttaacaatc ttctagctaa gaataatgtc gatacttatg gtatcagtct tgccagtcaa    4260
catggaacga cagatttgtt tagcacgctg gaacattacc gaaaagtctt tttaccaaat    4320
acaagcaata atgactggtt taaatcagag actaaggctt acattgtcga agaaaaatcc    4380
actatcgaag aggtgaaaac gaagcaaggg ttagctggca ccaagtattc tatcggtgtt    4440
tatgatcgta tcacgagtgc cacatggaaa taccgcaata tggtcttgcc tctcctgacc    4500
ttgccagaga gatccgtatt tgtcatctcg accatgtcta gtctaggatt tggagcttat    4560
gatcgctacc gcagtagtga ccataaagcg ggcaaggctc tcaatgattt tgttgaagaa    4620
aatgcgcgtg aaacagccaa acgtcagcga gatcactacg attattggta tcgtattttta    4680
gacgacaatg cacgtgaaaa actttataga aatattttgc ttttacgatgc ttataaattt    4740
ggcgatgata taccgtagg gaaagctaca gaagtggcag attttgataa tccaaatcct    4800
gcaatgcaac atttctttgg acctgttgga aataaagttg ggcataatca acacggtgct    4860
tatgctacag gtgatgcagt ttattatatg ggttatcgaa tgttggataa ggatggagct    4920
attacttata cgcatgagat gacacatgac tcagatcagg acatttatct tggaggatat    4980
ggtcgaagaa gtggcttggg accagagttc tttgctaaag gattattaca agcaccagac    5040
catccagatg atgcgaccat taccatcaac tccatcttga acattcaaa atctgatagt    5100
acagaaagtc gacgattaca agtacttgat ccaactacaa gatttaataa tgcagatgat    5160
ttgaagcaat atgtccacaa catgtttgac gttgtttata tgttggaata tctcgaagga    5220
aattcaattc ttaaattgga tacgaatcaa aaacaacaac ttcttagaaa agttacaaat    5280
gagtaccatc ctgatcctga tggaaataag gtctatgcaa caaatgttgt cagaaatcta    5340
acagtagaag aagttgaaag actacgttca ttcaatgatt tgattgataa taatattctt    5400
tcgtctaggg aatatgcctc aggtaaatac gaaagaaatg gctacttcac tattaagtta    5460
tttgcaccga tttatgctgc attaagtaat gatataggaa caccaggtga cctgatggga    5520
cgtcgtatag cctatgaact actagctgct aaaggcttta agatggtat ggtaccatat    5580
atctcaaacc aatacgaaga agaagccaaa caaaagggca agacaatcaa tctctacggt    5640
aaaacaagag gtttggttac agatgacttg gttttggaaa aggtatttaa taaccaatat    5700
catacttgga gtgagtttaa gaaagctatg tatcaagaac gacaagatca gtttgataga    5760
ttgaacaaag ttactttaa tgatacaaca cagccttggc aaacatttgc caagaaaact    5820
acaagcagtg tagatgaatt acagaaatta atggacgttg ctgttcgtaa ggatgcagaa    5880
cacaattact accattggaa taactacaat ccagacatag atagtgaagt ccacaagctc    5940
aagagagcaa tctttaaagc ctatcttgac caaacaaatg attttagaag ttcaattttt    6000
gagaataaaa aa                                                        6012
```

SeqID 49

```
atgaaaatta ataaaaaata tctagcaggt tcagtggcag tccttgccct aagtgtttgt        60
```

```
tcctatgagc ttggacgtta ccaagctggt caggataaga aaagagtctaa tcgagttgct    120
tatatagatg gtgatcaggc tggtcaaaag gcagaaaact tgacaccaga tgaagtcagt    180
aagagggagg ggatcaacgc cgaacaaatt gttatcaaga ttacggatca aggttatgtg    240
acctctcatg gagaccatta tcattactat aatggcaagg ttccttatga tgccatcatc    300
agtgaagagc tcctcatgaa agatccgaat tatcagttga aggattcaga cattgtcaat    360
gaaatcaagg gtggttatgt cattaaggta aacggtaaat actatgttta ccttaaggat    420
gcagctcatg cggataatat tcggacaaaa gaagagatta aacgtcagaa gcaggaacgc    480
agtcataatc ataactcaag agcagataat gctgttgctg cagccagagc ccaaggacgt    540
tatacaacgg atgatgggta tatcttcaat gcatctgata tcattgagga cacgggtgat    600
gcttatatcg ttcctcacgg cgaccattac cattacattc ctaagaatga gttatcagct    660
agcgagttag ctgctgcaga agcctattgg aatgggaagc agggatctcg tccttcttca    720
agttctagtt ataatgcaaa tccagctcaa ccaagattgt cagagaacca caatctgact    780
gtcactccaa cttatcatca aaatcaaggg gaaaacattt caagcctttt acgtgaattg    840
tatgctaaac ccttatcaga acgccatgtg gaatctgatg cccttatttt cgacccagcg    900
caaatcacaa gtcgaaccgc cagaggtgta gctgtccctc atggtaacca ttaccacttt    960
atcccttatg aacaaatgtc tgaattggaa aaacgaattg ctcgtattat tccccttcgt   1020
tatcgttcaa accattgggt accagattca agaccagaag aaccaagtcc acaaccgact   1080
ccagaaccta gtccaagtcc gcaaccagct ccaagcaatc caattgatga gaaattggtc   1140
aaagaagctg ttcgaaaagt aggcgatggt tatgtgcttg aggagatgg agtttctcgt    1200
tatatcccag ccaaggatct ttcagcagaa acagcagcag gcattgatag caaactggcc  1260
aagcaggaaa gtttatctca taagctagga actaagaaaa ctgacctccc atctagtgat  1320
cgagaatttt acaataaggc ttatgactta ctagcaagaa ttcaccaaga tttacttgat  1380
aataaaggtc gacaagttga ttttgaggct ttggataacc tgttggaacg actcaaggat  1440
gtctcaagtg ataaagtcaa gttagtggaa gatattcttg ccttcttagc tccgattcgt  1500
catccagaac gtttaggaaa accaaatgcg caaattacct acactgatga tgagattcaa  1560
gtagccaagt tggcaggcaa gtacacaaca gaagacggtt atatctttga tcctcgtgat  1620
ataaccagtg atgaggggga tgcctatgta actccacata tgacccatag ccactggatt  1680
aaaaaagata gtttgtctga agctgagaga gcggcagccc aggcttatgc taaagagaaa  1740
ggtttgaccc ctccttcgac agaccatcag gattcaggaa atactgaggc aaaaggagca  1800
gaagctatct acaaccgcgt gaaaagcagc aagaaggtgc cacttgatcg tatgccttac  1860
aatcttcaat atactgtaga agtcaaaaac ggtagtttaa tcatacctca ttatgaccat  1920
taccataaca tcaaatttga gtggtttgac gaaggccttt atgaggcacc taaggggtat  1980
actcttgagg atcttttggc gactgtcaag tactatgtcg aacatccaaa cgaacgtccg  2040
cattcagata atggttttgg taacgctagc gaccatgttc aaagaaacaa aaatggtcaa  2100
gctgatacca atcaaacgga aaaaccaagc gaggagaaac ctcagacaga aaaacctgag  2160
gaagaaaccc ctcgagaaga gaaaccgcaa agcgagaaac cagagtctcc aaaaaccaaca  2220
gaggaaccag aagaatcacc agagaatca gaagaacctc aggtcgagac tgaaaaggtt   2280
gaagaaaaac tgagagacga tgaagatta cttggaaaaa tccaggatcc aattatcaag  2340
tccaatgcca aagagactct cacaggatta aaaaataatt tactatttgg cacccaggac  2400
aacaatacta ttatggcaga agctgaaaaa ctattggctt tattaaagga gagtaag     2457

SeqID 50
ttgattttaa gtgtttgttc ttacgagttg ggactgtatc aagctagaac ggttaaggaa     60
aataatcgtg tttcctatat agatggaaaa caagcgacgc aaaaaacgga gaatttgact    120
cctgatgagg ttagcaagcg tgaaggaatc aatgctgagc aaatcgtcat caagataaca    180
gaccaaggct atgtcacttc acatgtcact cactatcatt attacaatgg taaggttcct    240
tatgacgcta tcatcagtga agaattactc atgaaagatc caaactataa gctaaaagat    300
gaggatattg ttaatgaggt caagggtgga tatgttatca aggtagatgg aaaaatactat   360
gtttaccttа aggatgctgc ccacgcggat aacgtccgta caaaagagga aatcaatcga    420
caaaaacaag agcatagtca acatcgtgaa ggtggaactc caagaaacga tggtgctgtt    480
gccttggcac gttcgcaagg acgctatact acagatgatg gttatatctt taatgcttct    540
gatatcatag aggatactgg tgatgcttat atcgttcctc atggagatca ttaccattac    600
attcctaaga atgagttatc agctagcgag ttggctgctg cagaagcctt cctatctggt    660
cgaggaaatc tgtcaaattc aagaacctat cgccgacaaa atagcgataa cacttcaaga    720
acaaactggg taccttctgt aagcaatcca ggaactacaa atactaacac aagcaacaca    780
agcaacactа acagtcaagc aagtcaaagt aatgacattg atagtctctt gaaacagctc    840
tacaaactgc ctttgagtca acgacatgta gaatctgatg ccttgtctt tgatccagca    900
caaatcacaa gtcgaacagc tagaggtgtt gcagtgccac acggagatca ttaccacttc    960
atcccttact ctcaaatgtc tgaattggaa gaacgaatcg ctcgtattat tccccttcgt   1020
tatcgttcaa accattgggt accagattca aggccagaac aaccaagtcc acaaccgact   1080
ccggaaccta gtccaggccc gcaacctgca ccaaatctta aatagactca aaattcttct   1140
ttggttagtc agctggtacg aaaagttggg gaaggatatg tattcgaaga aaagggcatc   1200
tctcgttatg tctttgcgaa agatttacca tctgaaactg ttaaaaatct tgaaagcaag   1260
ttatcaaaac aagagagtgt ttcacacact ttaactgcta aaaagaaaa tgttgctcct   1320
cgtgaccaag aatttttatga taaagcatat aatctgttaa ctgaggctca taagcccttg   1380
tttgaaaata agggtcgtaa ttctgatttc caagccttag acaaattatt agaacgcttg   1440
aatgatgaat cgactaataa agaaaaattg gtagatgatt tattggcatt cctagcacca   1500
attacccatc cagagcgact tggcaaacca aattctcaaa ttgagtatac tgaagacgaa   1560
```

```
gttcgtattg ctcaattagc tgataagtat acaacgtcag atggttacat ttttgatgaa   1620
catgatataa tcagtgatga aggagatgca tatgtaacgc ctcatatggg ccatagtcac   1680
tggattggaa aagatagcct ttctgataag gaaaaagttg cagctcaagc ctatactaaa   1740
gaaaaaggta tcctacctcc atctccagac gcagatgtta aagcaaatcc aactggagat   1800
agtgcagcag ctatttacaa tcgtgtgaaa ggggaaaaac gaattccact cgttcgactt   1860
ccatatatgg ttgagcatac agttgaggtt aaaaacggta atttgattat tcctcataag   1920
gatcattacc ataatattaa atttgcttgg tttgatgatc acacatacaa agctccaaat   1980
ggctatacct tggaagattt gtttgcgacg attaagtact acgtagaaca ccctgacgaa   2040
cgtccacatt ctaatgatgg atggggcaat gccagtgagc atgtgttagg caagaaagac   2100
cacagtgaag atccaaataa gaacttcaaa gcggatgaag agccagtaga ggaaacacct   2160
gctgagccag aagtccctca agtagagact gaaaaagtag aagcccaact caaagaagca   2220
gaagttttgc ttgcgaaagt aacggattct agtctgaaag ccaatgcaac agaaactcta   2280
gctggtttac gaaataattt gactcttcaa attatggata acaatagtat catggcagaa   2340
gcagaaaaat tacttgcgtt gttaaaagga agtaatcctt catctgtaag taaggaaaaa   2400
ataaac                                                            2406


SeqID 51
atgccagtag aaattaaaac cactaaagaa attcatccta aaatctatgc ctacaccaca     60
ccgacagtaa ccagtaatga aggctggatt aagattgggt atacagaacg tgatgtcaca    120
caacgtatca aggagcaaac gcatacagct catatagcta cagatgtctt atggactggt    180
gatgcagctt atacagaaga gcctgataag gggaaaactt tcaaggacca tgatttccac    240
catttccttt ctttccatga tgtagaacgt cgtcccaaga cggaatggtt ctattttaat    300
ggaactcctg aaaaatcaaa aaatcttttt gataagtttg ttcagcatga tttgtctggt    360
tatcagcctg gaaaaggaca ggactatact ctgcgacaag agcaagaaga agcagttgct    420
aagacattag cttatttcca agaacatgct ggaggcaagt ttctctgaga tgccaagcca    480
cgctttggta aaaccttgtc tacctatgac ctagctcgac ggatggaagc tgtcaatgtc    540
ctaattgtaa caaaccgccc tgccattgct aactcatggt atgatgattt tgaaacattc    600
atagcaggtc aaacgactta caagttgtt tctgaatcag atagccttaa gagtcgtcca    660
atcttgtcac gacaagaatt tcttggtatt ttagctgacg atgtaagaca acttgctttt    720
atcagtctcc aagacttgaa aggatctgtt tatttaggtg gagagcacga taaactcaaa    780
tgggtaactg atctgcattg ggacttgttg gttattgacg aggctcatga aggagttgat    840
accttcaaga ctgaccaagc ctttaataag attcgacgaa attttactct gcatttgtca    900
ggtacatcat ttaaagcatt ggctaaagga gattttacag aggaacaaat ctacaactgg    960
tcttatgctg atgagcaggc tgctaagtat tcgtggtctc ttgagcaaga agaggaaaat   1020
ccttatgaaa gcttgcctca gttgaatctc tttacctatc aaatgtctca gatgattggc   1080
gaaaagttag aaaaaggcgc tcagatcgat ggtgaaaata ttgactatgt ttttgactta   1140
agtgaatttt tcgctacaga tgataaaggg aaatttattc atgagcatga tgtcagaaat   1200
tggttagata ctctatcaag caatgaaaaa tatccatttt caaccaaaga actccgtaat   1260
gaactcaagc atactttttg gctttagaa cgtgtcgctt cggccaaagc attaaaagcc   1320
ctactagaag aacacccaat ctatgaaaac tatgagatcg ttctagctgc tggtgacgga   1380
cgtatgtccg aagaagacga taaagtcaaa ctcaaatcct tggacttggt tagaaaagcg   1440
atagcagaga atgacaaaac cattaccta tccgttggtc agctgacgac aggtgtcact   1500
atccctgaat ggacaggtgt attgatgtta tcaaatttga aatcaccagc tctttatatg   1560
caggccgcct tccgtgctca aaatcttac tcatggagcg ataacaaagg aaatcacttt   1620
cgcaaagaaa gagcctatgt atttgacttt gcgccggaaa gaaccttgat tctctttgat   1680
gagtttgcca acaacttatt gcttgtaact gcagctggta gaggaacttc agctacacgc   1740
gaagaaaata ttagagaatt attaaacttc tttccaatta ttgccgaaga ccgtgctggt   1800
aagatggttg aaattgatgc aaaggcagtt ctaaccactc ctcgccagat aaaagctaga   1860
gaagttctta aacgaggttt tatgtccaat ctcttatttg ataatattag tggtattttc   1920
caagcaagtc aaacagtttt agatatttta aatgagctgc cagttgaaaa ggaagggaag   1980
gtacaagata gttcgatttt attagatttct tcagtagtta cagtccgatga tgagggaaat   2040
gcagtagtag accatgaaat tgtagttaat cagcaaatgc gactttttgg tgaaaaagtt   2100
tatggacttg gtgaatctgt tgctgagtta gtcacaaaag atgaggaacg aactcaaaaa   2160
cagctggtca atgacttgag taagaccgtt tcttcagtga ttgtagagga attgaaagca   2220
gattattctc taaaaacaag ggaaactgag caaattaaga aacaaattac agcaacactt   2280
gagaatgaaa ttcgaaaaaa tgatatcgaa agaaaaattt ctgaagctca tatcaagcaa   2340
gagttgcaac agcagctcaa agaagcaaat gataaagcgc aaaaagataa gattcaagaa   2400
gatttggaaa aacgtttaga agaaataaa ctcattcata agaaaaact agaacaaaca   2460
ctcaaaaaag aagtggaaaa aatgcctgag aaatttatcg aacaggttga ataaaacgt   2520
gtggaacagt tgaaacaatc agctcaagat gaaattcgtg accatttacg agggtttgca   2580
agaacaattc caagttttat tatggcttac ggtgatcaaa ctctaacact tgataatttt   2640
gatgcctttg ttcctgaaca tgttttttat gaagtaacag ggattacgat tgatcagttt   2700
agatatttgc gagatggtgg gcaggatttt gcaggcatc tctttgataa agcaacattt   2760
gacgaagcta ttcaagaatt tcttcgcaag aaaaaggagt tggcggatta ttttaaagat   2820
caaaaagaag acattttgta ctatattcca ccgcagaaga ccaaccaaat tttcactcct   2880
aaacgagtgg tgaaaaggat ggtagatgat ttggaaaagg aaaatccagg gattttgat   2940
gatccatcta agactttat tgatttatat atgaagtcag gcctctatat tgcagaactt   3000
gtgaagcggt tatataatag caatggcttg aaagaggcct ttccaaatcc tgaagaacgc   3060
```

```
ttaaaacata ttttggaaaa gcaagtttat ggatttgctc cgtctgagat tatctataac    3120
atttccacta attttatatt tggcaatctt tctaaagata tcagtaggaa gaattttgtt    3180
ttagcagata ccattccagc ggctaaagaa gggagcattc aaaagttggt tgattcctat    3240
tttgaaaata at                                                        3252
```

SeqID 52
```
atgaaaaaaa tactaattgt agatgatgag aaaccaatct cggatattat caagtttaat      60
atgaccaagg aaggttacga agttgtaact gcttttaatg gtcgtgaagc gctagagcaa     120
tttgaagcag agcaaccaga tattattatt ctggatttga tgcttccaga aattgatggt     180
ttagaagttg ctaagaccat tcgtaagaca agcagtgtgc ccattcttat gctttcagcc     240
aaagatagtg aatttgataa ggttatcggt ttggaacttg gggcagatga ctatgtaaca     300
aaacccttct ccaatcgtga gttgcaggcg cgtgttaaag ctcttctgcg tcgttctcaa     360
cctatgccag tagatggtca ggaagcagat agtagcacca aacctatcca aattggggat     420
ttagaaattg ttccagacgc ctacgtggct aaaaaatatg gcgaagaact agacttaacc     480
catcgtgaat ttgagctttt gtatcattta gcatcgcata caggtcaagt catcacgcgc     540
gaacacttgc ttgagactgt ctggggttat gactatttg gtgatgtccg tacagttgat     600
gtgactgtac gacgtctgcg tgagaagatt gaagatacgc ccagccgacc agagtatatc     660
ttgacgcgcc gtggtgtagg gtattacatg agaaataatg ct                       702
```

SeqID 53
```
atgaagaaaa aatttctagc attttttgcta attttattcc caatttttctc attaggtatt     60
gccaaagcag aaacgattaa gattgtttct gataccgcct atgcaccttt tgagtttaaa     120
gattcagatc aaacttataa aggaattgat gttgacatta ttaacaaagt cgctgagatt     180
aaaggctgga acattcagat gtcctatcct ggatttgacg cagcagtcaa tgcggttcaa     240
gctgggcaag ccgacgctat catggcaggg atgacaaaga ctaaagaacg tgaaaaagtc     300
ttcaccatgt ctgatactta ctatgataca aaagttgtca ttgctactac aaagtcacac     360
aaaattagca agtacgacca attaactggc aaaaccgttg gtgttaaaaa cggaactgcc     420
gctcaacgtt tccttgaaac aatcaaagat aaatacggct ttactattaa aacatttgac     480
actggtgatt taatgaacaa cagcttgagt gctggtgcca tcgatgccat gatggatgac     540
aaacctgtta tcgaatatgc cattaaccaa ggtcaagacc tccatattga aatggatggt     600
gaagctgtag gaagttttgc tttcggtgtg aaaaaaggaa gtaaatacga gcacctggtt     660
actgaattca accaagcctt gtctgaaatg gtagtcttga taaaattatc     720
aagaaatgga ctgcttcatc atcttcagca gtgccaacta caactactct cgcaggatta     780
aaagctattc ctgttaaggc taaatatatc attgccagcg attcttcttt tgcccctttt     840
gttttccaaa attcaagcaa ccaatacact ggtattgata tggaattgat taaggcaatc     900
gctaaagacc aaggttttga aattgaaatc accaacctg gttttgatgc tgctatcagt     960
gctgtccaag ctggtcaagc cgatggtatc atcgctggta tgtctgtcac agatgctcgt    1020
aaggcaactt ttgacttctc agaatcatac tacactgcta ataccattct tggtgtcaaa    1080
gaatcaagca atattgcttc ttatgaagat ctaaaaggaa agacagtcgg tgttaaaaac    1140
ggaactgctt ctcaaacctt cctaacagaa aatcaaagca aatacggct caaaatcaaa    1200
acctttgcta tggttcttc aatgtatgac agtttaaaca ctggtgccat tgatgccgtt    1260
atggatgatg aacctgttct caaatattct atcagccaag tcaaaaatt gaaaactcca    1320
atctctggaa ctccaatcgg tgaaacagcc tttgccgtta aaaaggagc aaatccagaa    1380
ctgattgaaa tgttcaacaa cggacttgca aaccttaaag caaacggtga attccaaaag    1440
attcttgaca ataccctagc tagcgaatct tcaactgctt caacaagtac tgttgacgaa    1500
acaacgctct ggggcttgct tcaaaacaac tacaaacaac tccttagcgg tcttggtatc    1560
actcttgctc tagctcttat ctcatttgct attgccattg tcatcggaat tatcttcggt    1620
atgtttagcg ttagcccata caaatctctt cgcgtcatct ctgagatttt cgttgacgtt    1680
attcgtggta ttccattgat gattcttgca gccttcatct tctgggatac tccaaacttc    1740
atcgagtcta tcacaggcca acaaagccca attaacgact ttgtagctgg aaccattgcc    1800
ctctcactca atgcggctgc ttatatcgct gaaatcgttc gtggtggtat tcaggccgtt    1860
ccagttggcc aaatggaagc cagccgaagc ttgggtatct cttatggaaa aaccatgcgt    1920
aagattatct tgccacaagc aactaaattg atgttgccaa actttgtcaa ccaattcgtt    1980
atcgctctta agatacaac tatcgtatct gctatcggtt ggttgaact cttccaaact    2040
ggtaagatta tcattgctcg taactaccaa agtttcaaga tgtatgcaat ccttgctatc    2100
ttctatcttg taattatcac acttttgact agactagcga aacgcttaga aaagaggatt    2160
cgt                                                                  2163
```

SeqID 54
```
atggcatttg aaagtttaac agaacgtttg cagaacgtct ttaaaaatct acgtaaaaaa      60
ggaaaaatct ctgaatctga tgtccaagag caaccaaag aaattcgctt ggccttgctc     120
gaggccgacg ttgccttgcc tgttgtaaag gactttatca gaaaagttcg tgagcgtgca     180
gtcgggcatg aggtcattga tacacttaat cctgcgcaac agattattaa aatcgttgat     240
gaggaattga cagccgtttt aggttctgat acggcagaaa ttatcaagtc acctaagatt     300
ccaaccatca tcatgatggt tggtttacaa ggggctggta aaacaacctt gctggtaaa     360
ttggccaaca aactcaagaa agaagaaaat gctcgtcctt gatgattgc ggcggatatt     420
tatcgtccag ctgccattga ccagcttaag accttgggac aacagattga gtgcctgtc     480
tttgcacttg aaacagaagt accagctgtt gagattgtac gtcaaggttt ggagcaagcc     540
```

```
caaactaatc ataacgacta tgtcttgatt gatactgcgg gtcgtttgca gattgatgag    600
ctcctcatga atgagcttcg tgatgtgaaa gcattggctc aaccaaatga aatcttgctt    660
gtcgttgatg ctatgattgg tcaggaagca gccaatgttg cgcgtgagtt taatgctcag    720
ttggaagtga ctgggggtcat ccttaccaag attgatggcg atactcgtgg tggtgctgct    780
ctgtctgttc gtcacattac tggaaaacca atcaagttca ctggtacagg tgaaaagatt    840
acggacattg aaaccttcca cccagaccgc atgtctagcc gtatccttgg tatgggggat    900
atgctcactt tgattgagaa agcttctcag gaatacgatg aacaaaaagc ccttgaaatg    960
gctgagaaga tgcgcgaaaa caccttgat tttaatgatt tcatcgatca attagatcag   1020
gtgcaaaata tggggccgat ggaagacttg ctcaagatga ttccaggtat ggccaacaat   1080
ccagcccttc aaaacatgaa ggtggatgaa cgccagattg ctcgtaaacg tgccattgtg   1140
tcttcgatga cacctgaaga gcgtgaaaac ccagatttgt taaatccaag ccgtcgccgt   1200
cgtattgctg ctggttctgg aaatacattc gtcgaagtca ataaattcat caaggacttt   1260
aaccaggcta aacagctcat gcaggtgtt atgtctgggg atatgaataa aatgatgaag   1320
caaatggggga ttaatccaaa taaccttcct aaaaatatgc caaatatggg aggaatggat   1380
atgtctgccc ttgaaggaat gatgggacaa ggcggtatgc ctgacttatc agctctcgga   1440
ggagcaggaa tgccagatat gagccagatg tttggtggcg gtttgaaagg taaaattggt   1500
gaatttgcca tgaaacagtc catgaaacgt atggctaaca aaatgaagaa agcgaagaag   1560
aaacgcaag                                                           1569


SeqID 55
atgagccaaa tttggactaa agaaaaattt ataagccaag ttcaaggtgg agtcattgtt     60
tcttgtcaag ctttacctgg tgaagccctt tataatgaag aatttagctt gatgccttt    120
atggctaaag cagctttaga ggcaggagca gtgggcattc gcgcaaattc tgtgcgtgat    180
attaaagcaa ttcagaaagt agtagattta ccaataattg gaattatcaa aagggattat    240
ccacctcaag aaccatatat tactgctacg atgaaagaag tagatgaact tgtagaatgc    300
ggaacaacag tcattgcatt tgatgcaact ttaagaccaa gatatgatgg cttagttgtc    360
agtgaattta tcaaaaaaat aaaagaaaaa tatccgaatc aattgctgat ggcggatgta    420
agtaatttag atgaaggtct ctatgcattt aaatcaggcg ttgatttgt tggtacaaca    480
ttatcaggtt acacaagtac aagtgtacaa tcagatgagc ctgatttga actaatgaaa    540
aaattggctg attttaatat tccggtaatt gccgaaggaa aaattcatta tccagaacaa    600
ttaaaaaaag cttatagttt aggtgttacc agtgtagtca ttggtggagc gattacacgt    660
ccaaaagaaa ttgctcagcg atttattaat gtcatcaaa                          699


SeqID 56
atgagatatt taactgcagg agaatcacac ggcccccgtc taacagctat tattgaggga     60
attccagctg gacttccatt gacagctgag gatatcaatg aggaccttag acgccgtcag    120
ggtggctacg gtcgtggtgg tcgtatgaag attgagaatg accaggttgt ctttacttcg    180
ggcgttcgcc acgggaagac gacaggggcg cctattacta tggatgtcat caataaggac    240
caccagaaat ggctggacat catgtctgcg gaggacattg aagaccgcct taaaagcaag    300
cggaaaatta ctcatcctcg cccaggtcat gccgatttgg ttgggggggat taagtaccgt    360
tttgatgatt tgcgaaattc tttggagcgt tcatcagctc gtgaaaccac catgcgggtg    420
gcagttggtg cagtagccaa acgcctcttg gctgagctgg atatggagat tgccaaccat    480
gtcgtggtct ttggtggcaa ggaaatcgat gttcctgaaa atctgacagt cgctgaaatt    540
aagcaaaagag ctgcccagtc tgaagtttct attgtcaacc aagaacgaga acaggaaatc    600
aaggactata ttgaccaaat caaacgtgat ggtgatacca tcggtggggt tgtggagaca    660
gtcgtcggag gcgttccagt tggtcttggt tcctatgtcc aatgggatag aaaattggat    720
gcaagattgg ctcaagctgt tgtctctatc aatgccttta aaggggtgga atttggtctt    780
ggctttgagg ctggttatcg taaaggcagc caagttatgg atgaaattct ctggtctaaa    840
gaagacggtt atactcgccg taccaataat ctaggtggtt ttgaaggtgg tatgactaat    900
gggcaaccca tcgttgttcg tggggtcatg aaacccattc ctactcttta taaacctctt    960
atgagtgtgg atatcgaaac ccacgaacct tacaaggcaa ccgtggagag aagtgatccg   1020
actgctcttc cagctgcagg aatggtcatg gaagcagttg tagcaacggt tctggcgcaa   1080
gaaatcctcg aaaaattctc atcagataat cttgaggaac taaaagaagc ggtagccaaa   1140
caccgagact atacaaagaa ctat                                          1164


SeqID 57
atggtagtta tgaatagaat aagagtcagc aaaagggttg aaaagaagct tgctaagggg     60
ctagttttac tagaagccag tgatcttgag aatgtcaatc ttaaggatca ggaagtagag    120
gtgcagggtc aggaaggaaa ctttcttggg actgcctacc tttctcagca aaacaagggc    180
ttgggctggt ttatcagcaa agacaaggtg gccttcaatc aagctttctt tgaaacgttg    240
tttagaaaag ccaaagaaaa gagaaacgcc tactatcaag atgatttgac aactgcctt    300
cgtctcttca atcaagaggg agatggcttt gggggtctga cagttgacct ttatggcgac    360
tacgccgtct tttcttggta taactcttat gtttatcaga ttcgtcagac tatatcagaa    420
gcctttagac aggttttccc tgaggtttta ggagcttatg agaaaatccg ctttaagggt    480
ttggactatg aatctgccca tgtttatggt caagaagcac ctgacttttt caatgtttta    540
gaaaatggtg tcctgtatca agtctttatg aatgatggct tgatgacagg aattttccta    600
gaccagcatg aggttcgcgg tagtttagtt gacggcttgg ctatgggtaa atccttactc    660
aatatgtttt cctacacagc ggctttttca gtagctgcgg ccatgggagg agctagccat    720
```

```
acaacttctg ttgatctagc caaacgttca cgagaattgt ctcaagcgca ttttcaggca    780
aatgggctca gcacagacga gcatcgtttt atagtcatgg atgtctttga gtatttcaaa    840
tatgccaaac gcaaagactt gacctacgat gtgattgtcc tagatccgcc tagctttgct    900
cggaataaaa aacaaacttt ctctgtggcc aaggattatc acaagttgat ttcccagagt    960
cttgagattt taaatccggg agggattatc attgccagta ccaatgctgc caatgtttcc   1020
cgtcagaaat ttacagaaca aattgataaa ggttttgcag gaagaagtta ccagatttta   1080
aacaaatatg gtcttccagc agattttgcc tataataaaa aagatgaaag tagtaattac   1140
ctcaaggtga ttagtatgaa ggttagtaaa                                    1170
```

SeqID 58
```
atgacaaaaa cattaaaacg tcctgaggtt ttatcacctg cagggacttt agagaagcta     60
aaggtagctg ttcagtatgg agcagatgct gtctttatcg gtggtcaggc ctatggtctt    120
cgtagccgtg cgggaaactt tactttcgaa cagatggaag aaggcgtgca gtttgcggcc    180
aagtatggtg ccaaggtcta tgtagcggct aatatggtta tgcacgaagg aaatgaagct    240
ggtgctggtg agtggttccg taaactgcgt gatatcggga ttgcagcagt tatcgtatct    300
gacccagcct tgattatgat tgcagtgact gaagcaccag gccttgaaat ccacctttct    360
acccaagcca gtgccactaa ctatgaaacc cttgagttct ggaaagagct aggcttgact    420
cgtgtcgttt tagcgcgtga ggtttcaatg gaagaattag ctgagatccg caaacgtaca    480
gatgttgaaa ttgaagcctt tgtccatgga gctatgtgta tttcatactc tggacgttgt    540
actctttcaa accacatgag tatgcgtgat gccaaccgtg tgtggatgtt ctcagtcatgc    600
cgttggaaat acgaccttta cgatatgcca tttgggaaag aacgtaagag tttgcagggt    660
gagattccag aagaattttc aatgtcagcc gttgacatgt ctatgattga ccacattcca    720
gatatgattg aaaatggtgt ggacagtcta aaaatcgaag gacgtatgaa gtctattcac    780
tacgtatcaa cagtaaccaa ctgctacaag gcggctgtgg atgcctatct tgaaagtcct    840
gaaagtttg aagctatcaa acaagacttg gtggacgaga tgtggaaggt tgcccaacgt    900
gaactggcta caggatttta ctatggtaca ccatctgaaa atgagcagtt gtttggtgct    960
cgccgtaaaa ttcctgagta caagttcgtc gctgaagtgg tttcttatga tgatgcggca   1020
caaacagcaa caattcgtca acgaaatgtc attaacgaag gggaccaagt tgagttttat   1080
ggtccaggtt tccgtcattt tgaaacctat attgaagatt tgcatgatgc caaaggcaat   1140
aaaatcgacc gcgctccaaa tccaatggaa ctattgacta ttaaggtgcc tcaacccgtt   1200
caatcaggag atatggttcg tgcattaaaa gaaggactca tcaatcttta taaggaagat   1260
ggaaccagcg tcacagttcg agct                                          1284
```

SeqID 59
```
atgaatacct atcaattaaa taatggagta gaaattccag tattgggatt tggaactttt     60
aaggctaagg atggagaaga agcctatcgt gcagtgttag aagccttgaa ggctggttat    120
cgtcatattg atacggcggc gatttatcag aatgaagaaa gtgttggtca agcaatcaaa    180
gatagcggag ttccacgtga agaaatgttc gtaactacca agctttggaa tagtcagcaa    240
acctatgagc aaactcgtca agctttggaa aaatctatag aaaaactggg cttggattat    300
ttggatttgt atttgattca ttggccgaac ccaaaaccgc tcagagaaaa tgacgcatgg    360
aaaactcgca atgcggaagt ttggagagcg atggaagacc tctatcaaga agggaaaatc    420
cgtgctatcg gcgttagcaa tttttcttcc catcatttgg atgccttgct tgaaactgca    480
actatcgttc ctgcggtcaa tcaagttcgc ttggcgccag gtgtgtatcg agatcaagtc    540
gtagcttact gtcgtgaaaa gggaattta ttggaagctt gggggccttt tggacaagga    600
gaactgtttg atagcaagca agtccaagaa atagcagcaa atcacgtaaa atcggttgct    660
cagatagcct tggcctggag cttggcagaa ggattttac cacttccaaa atctgtcaca    720
acctctcgta ttcaagctaa tcttgattgc tttggaattg aactgagtca tgaggagaga    780
gaaaccttaa aaacgattgc tgttcaatcg ggtgctccac gagttgatga tgtggatttc    840
```

SeqID 60
```
ttgagtgaaa agtcaagaga agaagagaaa ttaagctta aagagcagat tctgagagat     60
ttagaaaaag taaaaggcta tgatgaagtt ctgaaagaag atgaggcagt agttcgcact    120
cctgcaaatg aaccttcaac tgaagaactc atggctgatt ccttgtcaac ggtagaggag    180
attatgagaa aagctcctac cgtgcctact caccaagtc aaggtgtacc agcttctcca    240
gcagatgaga ttcaaagaga aactcctggt gttccaagtc atccaagtca agatgtacct    300
tcttctccag cggaagaaag tggatcaaga ccaggtccag gtcctgttag acctaagaaa    360
cttgaaagag aatacaatga aaccccaaca agggtagctg tttcctatac gacggcagag    420
aaaaaagcag aacaagcagg tccagaaaca cctacgcctg ctacagaaac agtggatatc    480
atcagagata catcacgtcg tagccgtaga gaaggagcaa aacccgttaa gcctaagaaa    540
gagaagaagt cacatgtgaa agcttttgtg atttcattcc ttgtattcct tgccttgctc    600
tcagcaggtg gttactttgg ttaccagtac gtgctagatt ccttattacc tatcgatgct    660
aattctaaga aatatgtgac ggttggaatt ccagaaggtt caaacgttca agaaatcggt    720
acgacgcttg aaaaagctgg tttggtaaag catggtctga tttttagttt ttatgccaag    780
tataaaaatt ataccgactt gaaagcaggt tactacaatt gcaaaagag tatgagtaca    840
gaagacttac tcaaagagtt gcaaaaaggt ggaacagatg aaccgcaaga acctgtactt    900
gcgactttga caattccaga aggttatacc ttggatcaga ttgctcaagc tgtgggtcaa    960
ttgcaaggtg acttcaaaga gtctttgaca gcggaggctt cttggctaa agttcaagat   1020
gagacgttta tcagtcaagc agtagcgaaa tatcctactt tactggaaag tttgcctgta   1080
```

```
aaagacagcg gtgcgcgtta tcgtttggaa ggataccttt tcccagctac atactctatc   1140
aaggaaagca caactattga gagcttgatt gatgagatgt tagctgctat ggataagaac   1200
ctatctcctt actatagtac tatcaaatct aaaaacttga ctgtcaatga gttgttgacc   1260
attgcttcct tggtcgaaaa agaaggtgcc aagacagaag atcgtaagct cattgcaggt   1320
gtattctaca atcgtttgaa tcgtgatatg ccacttcaaa gtaatattgc aatcttgtat   1380
gcccaaggaa aactggggca aaatatcagt ctagctgagg atgttgcgat tgataccaac   1440
attgattcac cttataatgt ttataaaaat gtaggtctca tgcctggtcc agtcgatagt   1500
ccaagtctgg atgcgattga gtcaagcatc aatcaaacta agagcgataa cctctacttt   1560
gtagcagatg tcacagaagg caaggtctac tatgctaaca atcaagaaga ccacgaccgc   1620
aatgtcgctg aacatgtcaa cagcaaatta aac                                1653
```

SeqID 61

```
atgaaacaag aacgatttcc attggtgtca gatgacgagg tcatgttgac tgaaatgcca   60
gtcatgaatc tctatgatga gtctgatctg atcagtaata tcaagggtga gtatcgagat   120
aaaaattatt tagaatgggc tcctattgct gaagaaaaac cagtaaaacc gattgaaaag   180
caagtcgaaa aacctaaaaa ggctcctta ggggttaaaa aagaaggaaa gagcctatgcg   240
gaggtggcgc gtgaagaagc gcgtgcagaa ttgaaaaga aacgctctgc taactaccta   300
actcaggatt tcagccttgc gagacgtcat tctcagccca gtctagttag acagggcaat   360
caaccgacag ctcctttcca aaaggaaaat cctggtgaat ttgtcaaata tagccaaaaa   420
ttgacccagt ctcattatat cttggcggaa gaagttcatt ctatccctac caagaatgaa   480
gaagtgtcag cacctgctcc aaagaaaaac aattatgatt ttctaaagaa gagccaaatc   540
tacaataaaa aaagtaaaca aacagaacaa gaacgtcggg ttgcccaaga gttgaatctg   600
accagaatga cagaa                                                    615
```

SeqID 62

```
atgaaaaagt ctaagagcaa atatctaacc ttggcaggtc ttgtcctggg tacaggagtt   60
ttattgagcg cgtgtggaaa ttctagcacg gcgtcaaaaa cctacaacta tgtttattca   120
agtgatccat ctagcttgaa ctatctagca gaaaaccgcg cagcaacatc cgatattgtt   180
gcaaatttgg tagacgggtt attagaaaat gaccaatatg ggaatattat tccatcatta   240
gcagaggatt ggactgtttc tcaggacggt ttgacctata cctacaaact tcgtaaggat   300
gccaagtggt ttacttctga gggagaagaa tatgcgcctg taactgccca ggattttgtg   360
acaggtttgc aatatgcagc tgataaaaaa tcagaagcct tgtatctagt gcaggactct   420
gttgctggtt tggatgacta tatcactggt aaaacaagcg acttttcaac tgtcggtgtc   480
aaggcacttg atgaccaaac ggttcaatat actttggtta aaccagaact ttactggaat   540
tcaaaaacac ttgcaacgat acttttcct gttaatgcag atttcctgaa atcaaaaggg   600
gatgattttg ggaggcgga tccatctagt attttgtaca atggaccttt cttgatgaaa   660
gcacttgtct caaaatctgc tattgaatat aagaaaaacc ctaattactg ggatgctaag   720
aatgtctttg tagacgatgt gaaattgacc tactatgatg gtagcgacca agaatcactg   780
gaacgtaatt ttacagctgg tgcttatact acggctcgtc ttttcctaa cagctccagc   840
tatgaaggga ttaaagaaaa atacaaaaac aatatcatct atagtatgca aaattcaact   900
tcatatttct ttaatttaa cctagatagg aagtcttaca attatacttc taaaacaagt   960
gacattgaaa agaaatcgac tcaggaagca gttctcaata aaaacttccg tcaggctatc   1020
aattttgctt ttgacagaac atcttatggg gctcagtctg aagggaaaga aggtgcaaca   1080
aagatttttgc gtaacctagt ggttcctcca aactttgtca gtatcaaggg aaaagacttt   1140
ggtgaagttg tagcctctaa gatggtcaac tatggtaagg aatggcaagg tatcaacttt   1200
gcggatggtc aagacccta ctacaatcct gagaaagcca aggctaagtt tgcggaagct   1260
aagaaagaac tcgaagcaaa gggtgttcaa ttcccaatcc acttggataa gactgtggaa   1320
gtaacagata aagtaggcat acaaggagtt agttctatca acaatcaat tgaatctgtt   1380
ttaggttctg ataatgtagt gattgacatt cagcaattaa catcagatga gtttgacagt   1440
tcaggctact ttgctcaaac agctgctcag aaagattatg atttatatca tggcggttgg   1500
ggacctgatt atcaagaccc gtcaacctat ctcgatattt ttaatactaa tagtggagga   1560
tttctgcaaa atcttggact agagcctggt gaggccaatg acaaggctaa ggcagttgga   1620
ctggatgtct atactcaaat gttggaagaa gctaataaag agcaagatcc ggccaaacgt   1680
tatgagaaat atgctgatat tcaagcttgg ttgattgata gttcttagt tcttccaagt   1740
gtttcgcgtg ggggaacacc atcattgaga agaaccgtac catttgctgc tgcctatggt   1800
ttaaccggta caaaaggggt tgaatcatat aaatacctca agtacaaga taagattgtc   1860
acaacagacg aatatgcaaa agccagagaa aaatggttga agaaaaaga agaatccaat   1920
aaaaaagccc aagaagaatt ggcaaaacat gtcaaa                             1956
```

SeqID 63

```
gtggaacagc attcagatgt ctgttacatt ttttatagga gagaagatt gaaaacaaaa   60
attggattag caagtatctg tttactaggc ttggcaacta gtcatgtcgc tgcaaatgaa   120
actgaagtag caaaaacttc gcaggataca acgacagctt caagtagttc agagcaaaat   180
cagtcttcta ataaaacgca aacgagcgca gaagtacaga ctaatgctgc tgcccactgg   240
gatggggatt attatgtaaa ggatgatggt tctaaagctc aaagtgaatg gatttttgac   300
aactactata aggcttggtt ttatattaat tcagatggtc gttactcgca gaatgaatgg   360
catggaaatt actacctgaa atcaggtgga tatatggccc aaaacgagtg gatctatgac   420
agtaattaca agagttggtt ttatctcaag tcagatgggg cttatgctca tcaagaatgg   480
```

```
caattgattg gaaataagtg gtactacttc aagaagtggg gttacatggc taaaagccaa    540
tggcaaggaa gttatttctt gaatggtcaa ggagctatga tgcaaaatga atggctctat    600
gatccagcct attctgctta tttttatcta aaatccgatg gaacttatgc taaccaagag    660
tggcaaaaag tgggcggcaa atggtactat ttcaagaagt ggggctatat ggctcggaat    720
gagtggcaag gcaactacta tttgactgga agtggtgcca tggcgactga cgaagtgatt    780
atggatggta ctcgctatat cttttgcggcc tctggtgagc tcaaagaaaa aaaagatttg    840
aatgtcggct gggttcacag agatggtaag cgctatttct ttaataatag agaagaacaa    900
gtgggaaccg aacatgctaa gaaagtcatt gatattagtg agcacaatgg tcgtatcaat    960
gattggaaaa aggttattga tgagaacgaa gtggatggtg tcattgttcg tctaggttat   1020
agcggtaaag aagacaagga attggcgcat aacattaagg agttaaaccg tctgggaatt   1080
ccttatggtg tctatctcta tacctatgct gaaaatgaga ccgatgctga gagtgacgct   1140
aaacagacca ttgaacttat aaagaaatac aatatgaacc tgtcttaccc tatctattat   1200
gatgttgaga attgggaata tgtaaataag agcaagagag ctccaagtga tacaggcact   1260
tgggttaaaa tcatcaacaa gtacatggac acgatgaagc aggcgggtta tcaaaatgtg   1320
tatgtctata gctatcgtag tttattacag acgcgtttaa aacacccaga tattttaaaa   1380
catgtaaact gggtagcggc ctatacgaat gctttagaat gggaaaaccc tcattattca   1440
ggaaaaaaag gttggcaata tacctcttct gaatacatga aaggaatcca agggcgcgta   1500
gatgtcagcg tttgggtat                                                1518
```

SeqID 64
```
atggcaaaag aaccgtggca agaagatatc tatgatcaag aagaatcaag agcagagcgt     60
cggcatcgaa accacggagg ggctgatagg atggctaatc gtattttgac gatcctagct    120
agtattttct ttgtaattgt ggtggtgatg gtcatcgttc tcatctatct atcatcgggg    180
gggagtaatc gcacagcagc cttaaaaggc tttcatgatt ctgatgccag tgtagtacaa    240
atctcatctt caagtagttc tcagcctgag cagagttcag agccagaatc tacttctagt    300
agttcagaag aagctgctaa tcctgaagga acgattaaag ttctcgcagg agaaggggaa    360
gcagctattg ccgctcgtgc aggaatctcc attgctcagt tagaggcctt gaatcctggg    420
cacatggcta caggatcttg gtttgctaat ccaggtgatg ttataaaaat aaaa          474
```

SeqID 65
```
atgccaatta catcattaga aataaaggac aagacttttg gaactcgatt cagaggtttt     60
gatccagaag aagtcgatga attttttagat attgtggttc gtgattacga agatcttgtg    120
cgtgcgaatc atgataaaaa tttgcgtatt aagagtttag aagagcgttt gtcttacttt    180
gatgaaataa aagattcatt gagccagtct gtattgattg ctcaggatac agctgagaga    240
gtgaaacagg cggcgcataa acgttcaaac aatatcattc atcaagcaga gcaagatgcg    300
caacgcttgt tggaagaagc taaatataag gcaaacgaga ttcttcgtca agcaactgat    360
aatgctaaga aagtcgctgt tgaaacagaa gaattgaaga caagagccg tgtcttccac     420
caacgtctca aatctacaat tgagagtcag ttggctattg ttgaatcttc agattgggaa    480
gatattctcc gtccaacagc tacttatctt caaaccagtg atgaagcctt taaagaagtg    540
gttagcgaag tacttggaga accgattcca gctccaattg aagaagaacc aattgatatg    600
acacgtcagt tctctcaagc agaaatggca gaattacaag ctcgtattga ggtagccgat    660
aaagaattgt ctgaatttga agctcagatt aaacaggaag tggaagctcc aactcctgta    720
gtgagtcctc aagttgaaga agagcctctg ctcatccagt tggcccaatg tatgaagaac    780
cagaag                                                               786
```

SeqID 66
```
atgtctttaa aagatagatt cgatagattt atagattatt ttacggagga tgaggattca     60
agtctccctt atgaaaaaag agatgagcct gtgtttactt cagtaaattc ttcacaggaa    120
ccggctctcc caatgaatca accttacag tcggctggca caaaagagaa caatatcacc      180
agacttcatg caagacaaca ggaattggca aatcagagtc agcgtgcaac ggataaggtc    240
attatagatg ttcgttatcc tagaaaatat gaggatgcaa cagaaattgt tgatttattg    300
gcaggaaacg aaagtatctt gattgatttt cagtatatga cagaggtgca ggctcgtcgt    360
tgtttggact atttggatgg agcttgtcat gtttttagctg gaaatttgaa aaaggtagct    420
tctaccatgt atttgttgac accagtgaac gttattgtaa atgttgaaga tatccgttta    480
ccagatgaag atcaacaggg tgagttcggt tttgatatga agcgaaatag agtacga       537
```

SeqID 67
```
atgtcagatt tgaaaaaata cgaaggtgtc attccagcct tctacgcatg ttatgatgat     60
caaggagaag taagcccaga acgtacgcgt gccttggttc aatacttcat tgataaaggt    120
gttcaaggtc tttatgtcaa tggttcttct ggtgaatgta tctaccaaag cgttgaagat    180
cgcaagttca tttttggagga agtcatggcg gtagccaaag gtaaattgac cattattgcc    240
catgttgctt gcaataatac taaagatagt atggaacttg ctcgccatgc tgaaagcttg    300
ggagtagatg ctattgcaac gattccacca atttattttcc gcttgccaga atactcagtt    360
gccaaatact ggaacgtatc agttctgca gctccaaaca cagactacgt gatttacaac     420
attcctcaat ggcagggg tgctttgact ccaagccttt acacagaaat gttgaaaaat      480
cctcgtgtta tcggtgtgaa gaactcttct atgccagttc aagatatcca aaccttgtc      540
agccttggtg gagaagacca tatcgtcttt aatggtcctg atgagcagtt cctaggagga    600
cgcctcatgg gggctagggc tggtatcggt ggtacttatg gtgctatgcc agaactcttc    660
```

```
ttgaaactca atcagttgat tgcggataag gacctagaaa cagcgcgtga attgcagtat      720
gctatcaacg caatcattgg taaactcact tctgctcatg gaaatatgta cggtgtcatc      780
aaagaagtct tgaaaatcaa tgaaggcttg aatattggat ctgttcgttc accattgaca      840
ccagtgactg aagaagatcg tccagttgta gaagcggctg ctgccttgat tcgtgaaacc      900
aaggagcgct tcctc                                                      915


SeqID 68
atgaataaaa gaggtctttta ttcaaaacta ggaatttccg ttgtaggcat tagtctttta      60
atgggagtcc ccactttgat tcatgcgaat gaattaaact atggtcaact gtccatatct     120
cctatttttc aaggaggttc atatcaactg aacaataaga gtatagatat cagctctttg     180
ttattagata aattgtctgg agagagtcag acagtagtaa tgaaatttaa agcagataaa     240
ccaaactctc ttcaagcttt gtttggccta tctaatagta aagcaggctt taaaaataat     300
tactttttcaa ttttcatgag agattctggt gagataggtg tagaaataag agacgcccaa     360
aagggaataa attatttatt ttccagacca gcttcattat ggggaaaaca taaaggacag     420
gcagttgaaa atacactagt atttgtatct gattctaaag ataaaacata cacaatgtat     480
gttaatggaa tagaagtgtt ctctgaaaca gttgatacat ttttgccaat ttcaaatata     540
aatggtatag ataaggcaac actaggagct gttaatgtg aaggtaagga acattacctc     600
gcaaaaggaa gtattgatga aatcagtcta tttaacaaag caattagtga tcaggaagtt     660
tcaactattc ccttgtcaaa tccatttcag ttaatttcc aatcaggaga ttctactcaa     720
gctaactatt ttagaatacc gacactatat acattaagta gtggaagagt tctatcaagt     780
attgatgcac gttatggtgg gactcatgat tctaaaagta agattaatat tgccacttct     840
tatagtgatg ataatgggaa aacgtggagt gagccaattt ttgctatgaa gtttaatgac     900
tatgaggagc agttagttta ctggccacga gataataaat taaagaatag tcaaattagt     960
ggaagtgctt cattcataga ttcatccatt gttgaagata aaaaatctgg gaaaacgata    1020
ttactagctg atgttatgcc tgcgggtatt ggaaataata atgcaaataa agccgactca    1080
ggttttaaag aaataaatgg tcattattat ttaaaactaa agaagaatgg agataacgat    1140
ttccgttata cagttagaga aaatggtgtc gtttataatg aaacaactaa taaacctaca    1200
aattatacta taaatgataa gtatgaagtt ttggagggag gaaagtcttt aacagtcgaa    1260
caatattcgg ttgattttga tagtggctct ttaagagaaa ggcataatgg aaaaacaggtt    1320
cctatgaatg ttttctacaa agattcgtta tttaaagtga ctcctactaa ttatatagca    1380
atgacaacta gtcagaatag aggagagagt tgggaacaat ttaagttgtt gcctccgttc    1440
ttaggagaaa aacataatgg aacttactta tgtcccggac aaggtttagc attaaaatca    1500
agtaacagat tgattttgc aacatatact agtggagaac taacctatct catttctgat    1560
gatagtggtc aaacatggaa gaaatcctca gcttcaattc cgtttaaaaa tgcaacagca    1620
gaagcacaaa tggttgaact gagagatggt gtgattgata cattctttag aaccactaca    1680
ggtaagatag cttatatgac tagtagagat tctggagaaa catggtcgaa agtttcgtat    1740
attgatggaa tccaacaaac ttcatatggc acacaagtat ctgcaattaa atactctcaa    1800
ttaattgatg gaaaagaagc agtcattttg agtacaccaa attctagaag tggccgcaag    1860
ggaggccaat tagttgtcgg tttagtcaat aaagaagatg atagtattga ttggaaatac    1920
cactatgata ttgatttgcc ttcgtatggt tatgcctatt ctgcgattac agaattgcca    1980
aatcatcaca taggtgtact gtttgaaaaa tatgattcgt ggtcgagaaa tgaattgcat    2040
ttaagcaatg tagttcagta tatagatttg gaaattaatg atttaacaaa a            2091


SeqID 69
atgaatcgga gtgttcaaga acgtaagtgt cgttatagca ttaggaaact atcggtagga      60
gcggtttcta tgattgtagg agcagtggta tttggaacgt ctcctgtttt agctcaagaa     120
ggggcaagtg agcaacctct ggcaaatgaa actcaacttt cgggggagag ctcaaccca     180
actgatacag aaaagagcca gccttcttca gagactgaac tttctggcaa taagcaagaa     240
caagaaagga aagataagca agaagaaaaa attccaagag attactatgc acgagatttg     300
gaaaatgtcg aaacagtgat agaaaagaa gatgttgaaa ccaatgcttc aaatggtcag     360
agagttgatt tatcaagtga actagataaa ctaaagaaac ttgaaaacgc aacagttcac     420
atggagttta agccagatgc caaggcccca gcattctata atctcttttc tgtgtcaagt     480
gctactaaaa aagatgagta cttcactatg gcagtttaca ataatactgc tactctagag     540
gggcgtggtt cggatgggaa acagttttac aataattaca acgatgcacc cttaaaagtt     600
aaaccaggtc agtggaattc tgtgactttc acagttgaaa aaccgacagc agaactacct     660
aaaggccgag tgcgcctcta cgtaaacggg gtattatctc gaacaagtct gagatctggc     720
aatttcatta agatatgcc agatgtaacg catgtgcaaa tcggagcaac caagcgtgcc     780
aacaatacgg tttggggggtc aaatctacag attcggaatc tcactgtgta taatcgtgct     840
ttaacaccag aagaggtaca aaaacgtagt caacttttta acgctcaga tttagaaaaa     900
aaactacctg aaggagcggc tttaacagag aaaacggaca tattcgaaag cgggcgtaac     960
ggtaacccaa ataaagatgg aatcaagagt tatcgtattc cagcacttct caagacagat    1020
aaaggaactt tgatcgcagg tgcagatgca cgccgtctcc attcggagtga ctggggtgat    1080
atcggtatgg tcatcagacg tagtgaagat aatggtaaaa cttggggtga ccgagtaacc    1140
attaccaact tacgtgacaa tccaaaagct tctgacccat cgatcggttc accagtgaat    1200
atcgatatgg tgttggttca agatcctgaa accaaacgaa tcttttctat ctatgacatg    1260
ttcccagaag ggaagggaat ctttggaatg tcttcacaaa aagaagaagc ctacaaaaaa    1320
atcgatggaa aaacctatca aatcctctac cgtgaaggag aaaaggggagc ttataccatt    1380
cgagaaaatg gtactgtcta tacaccagat ggtaaggcga cagactatcg cgttgttgta    1440
```

```
gatcctgtta aaccagccta tagcgacaag ggtgatctat acaagggtga ccaattacta    1500
ggaaatatct acttcacaac aaacaaaact tctccattta gaattgccaa ggatagctat    1560
ctatggatgt cctacagtga tgacgacggg aagacatggt cagctcctca agatattact    1620
ccgatggtca aagccgattg gatgaaattc ttgggtgtag gtcctggaac aggaattgta    1680
cttcggaatg ggcctcacaa gggacggatt ttgataccgg tttatacgac taataatgta    1740
tctcacttag atggctcgca atcttctcgt gtcatctatt cagatgatca tggaaaaact    1800
tggcatgctg gagaagcggt caacgataac cgtcaggtag acggtcaaaa gatccactct    1860
tctacgatga acaatagacg tgcgcaaaat acagaatcaa cggtggtaca actaaacaat    1920
ggagatgtta aactctttat gcgtggtttg actggagatc ttcaggttgc tacaagtaaa    1980
gacggaggag tgacttggga gaaggatatc aaacgttatc cacaggttaa agatgtctat    2040
gttcaaatgt ctgctatcca tacgatgcac gaaggaaaag aatacatcat cctcagtaat    2100
gcaggtggac cgaaacgtga aaatgggatg gtccacttgg cacgtgtcga agaaaatggt    2160
gagttgactt ggctcaaaca caatccaatt caaaaaggag agtttgccta taattcgctc    2220
caagaattag gaaatgggga gtatggcatc ttgtatgaac atactgaaaa aggacaaaat    2280
gcctataccc tatcatttag aaaatttaat tgggaatttt tgagcaaaaa tctgatttct    2340
cctaccgaag cgaactagag agatgggcaa aggagagatg ggcaaaggag ttattggctt    2400
ggagttcgac tcagaagtat tggtcaacaa ggctccaacc cttcaattgg caaatggtaa    2460
aacagcgact ttcctaaccc agtatgatag caagaccttg ttgtttgcag tagataagga    2520
agatatcgga caggaaatta ttggtatagc taaaggaagc atcgaaagta tgcataatct    2580
tcctgtaaat ctagcaggtg ccagagttcc tggcggagta aatggtagca aagcagcggt    2640
gcatgaagtt ccagaattta caggggggagt taatggtaca gagccagctg ttcatgaaat    2700
cgcagagtat aagggatctg attcgcttgt aactcttact acaaaaaaag attatactta    2760
caaagctcct cttgctcagc aggcacttcc tgaaacagga aacaaggaga gtgacctcct    2820
agcttcacta ggactaacag ctttcttcct tggtctgttt acgctaggga aaagagaga    2880
acaa                                                                 2884
```

SeqID 70
```
atgatccaaa tcggcaagat ttttgccgga cgctatcgga ttgtcaaaca gattggtcga     60
ggaggtatgg cggatgtcta cctagccaaa gacttaatct tagatgggga agaagtggca    120
gtgaaggttc tgaggaccaa ctaccagacg gacccgatag ctgtagctcg ttttcagcgt    180
gaagcgagag ctatggcaga tctagaccat cctcatatcg ttcggataac agatattggc    240
gaggaagacg gtcaacagta cctagctatg gagtatgtgg ctggactgga cctcaaacgc    300
tatatcaagg aacattatcc tctttctaat gaagaagcag tccgtatcat gggacaaatt    360
ctcttggcta tgcgcttggc ccatactcga ggaattgttc acagggactt gaaacctcaa    420
aatatcctct tgacaccaga tgggactgcc aaggtcacag actttgggat tgctgtagcc    480
tttgcagaga caagtctgac ccagactaac tcgatgttga gctcagttca ttacttgtca    540
ccagagcagg cgcgtggttc gaaggcgact gtgcagagtg atatctatgc catggggatt    600
attttctatg agatgctgac aggccatatc ccttatgacg gggatagcgc ggtgaccatt    660
gccctccagc atttccagaa acccctgccg tccgttattg cagaaaatcc atctgtacct    720
caggctttag aaaatgttat tatcaaggca actgctaaaa agttgaccaa tcgctaccgc    780
tcggtttcag agatgtatgt ggacttgtct agtagcttgt cctacaatcg tagaaatgaa    840
agtaagttaa tctttgatga aacgagcaag gcagatacca agaccttgcc gaaggtttct    900
cagagtacct tgacatctat tcctaaggtt caagcgcaaa cagaacacaa atcaatcaaa    960
aacccaagcc aggctgtgac agaggaaact taccaaccac aagcaccgaa aaaacataga   1020
tttaagatgc gttacctgat tttgttggcc agccttgtat tggtggcagc ttctcttatt   1080
tggatactat ccagaactcc tgcaaccatt gccattccag atgtggcagg tcagacagtt   1140
gcagaggcca aggcaacgct caaaaaagcc aatttttgaga ttggtgagga agaagacagag   1200
gctagtgaaa aggtggaaga agggcggatt atccgtacag atcctggcgc tggaactggt   1260
cgaaaagaag gaacgaaaat caatttggtt gtctcatcag gcaagcaatc tttccaaatt   1320
agtaattatg tcggtcggaa atcctctgat gtcattgcgg aattaaaaga gaaaaaagtt   1380
ccagataatt tgattaaaat tgaggaagaa gagtcgaatg agagtgaggc tggaacggtc   1440
ctgaagcaaa gtctaccaga aggtacgacc tatgacttga gcaaggcaac tcaaattgtt   1500
ttgacagtag ctaaaaaagc tacgacgatt caattaggga actatatggg acggaactct   1560
acagaagtaa tctcagaact caagcgagag aaggttcctg agaatttgat taagatagag   1620
gaagaagagt ccagcgaaag cgaaccagga acgattatga aacaaagtcc aggtgccgga   1680
acgacttatg atgtgagtaa acctactcaa attgtcttga cagtagctaa aaaagttaca   1740
agtgttgcca tgccgagtta cattggttct agcttggagt ttactaagaa caatttgatt   1800
caaattgttg ggattaagga agctaatata gaagttgtag aagtgacgac agcgcctgca   1860
ggtagtgcag aaggcatggt tgttgaacaa agtcctagag caggtgaaaa ggtagacctc   1920
aataagacta gagtcaagat ttcaatctac aaacctaaaa caacttcagc tactcct      1977
```

SeqID 71
```
atgacaaaac taatctttat ggggacccccc gactttttcag caacagtctt aaaaggactt     60
ttgacagatg accgttacga aattctagcc gttgtgaccc agccagaccg tgctgttggt    120
cgtaaaaaag ttatccaaga aaccccagtc aagcaggctg ccaaggaagc aggactatct    180
atctaccaac ctgaaaaatt atctggaagt ccagagatgg aagatcttat gaagctagga    240
gcagatggaa ttgtgactgc tgctttttggg cagtttctcc caagcaaact ccttgatagc    300
atggactttg ctgtcaacgt tcatgcctcc ctccttccta gacaccgtgg tggtgcgcct    360
```

```
atccattatg ccttgattca aggggatgag gaagctggtg tgaccatcat ggaaatggtt      420
aaggaaatgg atgcaggaga tatgatttct cgtcgcagca ttccgatcac agatgaggac      480
aatgttggca ccttgtttga aaaattggcg ctagttggtc gtgatttgct tttggacact      540
ctgcctgcct atattgctgg tgatatcaaa cctgaaccgc aggatacgag tcaggttacc      600
ttctctccaa atataaagcc agaggaagaa aaactggact ggaacaaaac caatcgtcaa      660
ctctttaacc aaaattcgtgg aatgaacccc tggcctgttg cccatacttt ccttaagggc      720
gaccgcttta agatttatga agccctacca gtagaaggtc agggaaatcc aggtgagatt      780
ctctctatcg gcaagaaaga attgattgtc gcaacggctg aaggggctct atccctcaaa      840
caagtgcagc cagctggtaa gcctaagatg gacattgctt ccttcctcaa cggagttgga      900
cgtacattga ctgtaggaga acgatttggt gac                                   933


SeqID 72
gtgtttagac gtttaggtca agatttccag cttaggaaag tgaaaaagat tttaaagcag       60
attaatgccc tgaaaggcaa gatgtcctct ctttcggatc aagaattagt agctaaaaca      120
gtagagtttc gtcagcgtct ttccgaggga gaaagtctag acgatatttt ggttgaagct      180
tttgctgtgg tgcgtgaagc agataagcgg attttaggga tgtttcctta tgatgttcaa      240
gtcatgggag ctattgtcat gcactatgga aatgttgctg agatgaatac gggggaaggt      300
aagaccttga cagctaccat gcctgtctat ttgaacgctt tttcaggaga aggagtgatg      360
gttgtgactc ctaatgagta tttatcaaag cgtgatgccg aggaaatggg tcaagtttat      420
cgttttctag gattgaccat tggtgtacca tttacggaag atccaaagaa ggagatgaaa      480
gctgaagaaa agaagcttat ctatgcttcg gatatcatct acacaaccaa tagtaattta      540
ggttttgatt atctaaatga taacctagcc tcgaatgaag aaggtaagtt tttacgaccg      600
tttaactatg tgattattga tgaaattgat gatatcttgc ttgatagtgc acaaactcct      660
ctgattattg cgggttctcc tcgtgttcag tctaattact atgcgatcat tgatacactt      720
gtaacaacct tggtcgaagg agaggattat atctttaaag aggagaaaga ggaggtttgg      780
ctcactacta aggggccaa gtctgctgag aatttcctag ggattgataa tttatacaag      840
gaagagcatg cgtcttttgc tcgtcatttg gtttatgcga ttcgagctca taagctcttt      900
actaaagata aggactatat cattcgtgga aatgagatgg tactggttga taagggaaca      960
gggcgtctaa tggaaatgac taaacttcaa ggaggtctcc atcaggctat tgaagccaag     1020
gaacatgtca aattatctcc tgagacgcgg gctatggcct cgatcaccta tcagagtctt     1080
tttaagatgt ttaataagat atctggtatg acagggacag gtaaggtcgc ggaaaaagag     1140
tttattgaaa cttacaatat gtctgtagta cgcattccaa ccaatcgtcc gagacaacgg     1200
attgactatc cagataatct atatatcact ttacctgaaa aagtgtatgc atccttggag     1260
tacatcaagc aataccatgc taagggaaat cctttactcg tttttgtagg ctcagttgaa     1320
atgtctcaac tctattcgtc tctcttgttt cgtgaaggga ttgcccataa tgtcctaaat     1380
gctaataatg cggcgcgtga ggctcagatt atctccgagt caggtcagat gggggctgtg     1440
acagtggcta cctctatggc aggacgtggt acggatatca agcttggtaa aggagtcgca     1500
gagcttgggg gcttgattgt tattgggact gagcggatgg aaagtcagcg gatcgaccta     1560
caaattcgtg gccgttctgg tcgtcaggga gatcctggta tgagtaaatt ttttgtatcc     1620
ttagaggatg atgttatcaa gaaatttggt ccatcttggg tgcataaaaa gtacaaagac     1680
tatcaggttc aagatatgac tcaaccggaa gtattgaaag gtcgtaaata ccggaaacta     1740
gtcgaaaagg ctcagcatgc cagtgatagt gctggacgtt cagcacgtcg tcagactctg     1800
gagtatgctg aaagtatgaa tatacaacgg gatatagtct ataaagagag aaatcgtcta     1860
atagatggtt ctcgtgactt agaggatgtt gttgtggata tcattgagag atatacagaa     1920
gaggtagcgg ctgatcacta tgctagtcgt gaattattgt ttcacttttat tgtgaccaat     1980
attagttttc atgttaaaga ggttccagat tatatagatg taactgacaa aactgcagtt     2040
cgtagcttta tgaagcaggt gattgataaa gaactttctg aaaagaaaga attacttaat     2100
caacatgact tatatgaaca gttttttacga ctttcactgc ttaaagccat tgatgacaac     2160
tgggtagagc aggtagacta tctacaacag ctatccatgg ctatcggtgg tcaatctgct     2220
agtcagaaaa atccaatcgt agagtactat caagaagcct acgcgggctt tgaagctatg     2280
aaagaacaga ttcatgcgga tatggtgcgt aatctcctga tggggctggt tgaggtcact     2340
ccaaaaggtg aaatcgtgac tcatttttcca                                      2370


SeqID 73
atgaccgaaa cggtagaaga taaagtaagt cattcaatta ctgggcttga tatcctcaag       60
gggatagttg ctgcgggagc tgtcataagt ggaaccgttg caactcaaac gaaggtattt      120
acaaatgagt cagcagtact tgaaaaaact gtagagaaaa cggatgcttt ggcaacaaat      180
gatacagtag ttctaggtac gatatctaca agtaattcag cgagttcaac tagtttgtca      240
gcttcagagt cggcaagtac atctgcatct gagtcagcct caaccagcgc ttcgacctca      300
gcaagtacaa gtgcatcaga atcagcaagt acatcggctt cgacaagtat ttctgcatca      360
tctactgtgg taggttcaca aacagctgcc gctacagaag caactgctaa gaaggtcgaa      420
gaagatcgta agaaaccagc tagtgattat gtagcatcag ttacaaatgt caatctccaa      480
tcttatgcta agcgacgcaa gcgttcagtg gattccatcg agcaattgct ggcttctata      540
aaaaatgctg ctgttttttc tggcaatacg attgtaaatg gcgcccctgc aattaatgca      600
agtctaaaca ttgctaaaag tgagacaaaa gtttatacag gtgaaggtgt agattcggta      660
tatcgtgttc caatttacta taaattgaaa gtgacaaatg atggttcaaa attgacCttt      720
acctatacgg ttacgtatgt gaatcctaaa acaaatgatc ttggtaatat atcaagtatg      780
cgtcctggat attctatcta taattcaggt acttcaacac aaacaatgtt aacccttggc      840
```

```
agtgatcttg gtaaaccttc aggtgtaaag aactacatta ctgacaaaaa tggtagacag    900
gttctatcct ataatacatc tacaatgacg acgcagggta gtgggtatac ttggggaaat    960
ggtgcccaaa tgaatggttt ctttgctaag aaaggatatg gattaacatc atcttggact   1020
gtaccaatta ctggaacgga tacatccttt acatttaccc cttacgctgc tagaacagat   1080
agaattggaa ttaactactt caatggtgga ggaaaggtag ttgaatctag cacgaccagt   1140
cagtcacttt cacagtctaa gtcactctca gtaagtgcta gtcaaagcgc ctcagcttca   1200
gcatcaacaa gtgcgtcggc ttcagcatca accagtgcct cggcttcagc gtcaaccagt   1260
gcgtcagctt cagcaagtac cagtgcttca gtctcagcat caacaagtgc ttcagcctca   1320
gcatcgacaa gtgccteggc ttcagcaagc acatcagcat ctgaatcagc gtcaaccagt   1380
gcttcggctt cagcaagtac cagtgcttca gcttcagcat caaccagcgc ctcggcctca   1440
gcaagcacct cagcttctga atcggcctca accagcgcct cggcctcagc aagcacctca   1500
gcttctgaat cggcctcaac cagcgcctca gcctcagcat caacgagtgc ttcggcttca   1560
gcaagcacaa gcgcctcggg ttcagcatca acgagtacgt cagcttcagc gtcaaccagt   1620
gcttcagcct cagcatcaac aagtgcgtca gcctcagcaa gtatctcagc gtctgaatcg   1680
gcatcaacga gtgcgtctga gtcagcatca acgagtacgt cagcctcagc aagcacctca   1740
gcttctgaat cggcctcaac cagtgcgtca gcctcagcat cgacaagcgc ctcagcttca   1800
gcaagtacca gtgcttcaga ctcagcgtcg acaagtgcgt cggcctcaac cagtgcatct   1860
gaatcggcat caaccagtgc gtcagcctca gcaagtacta gtgcatcggc ttcagcatca   1920
accagtgcct cggcttcagc gtcaaccagt gcgtcagctt cagcaagtac cagtgcttca   1980
gtctcagcat caacaagtgc ttcagcctca gcatcgacaa gtgcctcggc ttcagcaagc   2040
acatcagcat ctgaatcagc gtcgacaagc gcctcagctt cagcaagtac cagtgcgtca   2100
gcctcagcgt cgacaagtgc gtcagcctca gcaagtacta gtgcatcagc ttcagcatca   2160
acgagtgcat cggcttcggc gtcaaccagt gcatcagagt cagcaagtac cagtgcgtca   2220
gcttccgcat caacaagtgc ctcggcttca gcaagcacca gtgcgtcggc ttcagcaagt   2280
actagcgcct cagcctcagc ctcaaccagt gcgtcagcct cagcaagtat ctcagcgtct   2340
gaatcggcat caacgagtgc gtccgcttca gcaagtacta gcgcctcagc ctcagcgtca   2400
acaagtgcat cggcttcagc gtcaacgagt gcgtctgaat cggcatcaac gagtgcgtcc   2460
gcttcagcaa gtactagcgc ctcagcctca gcgtcaacaa gtgcatcggc ttcagcatca   2520
acgagtgcgt ccgcttcagc aagtactagc gcctcagcct cagcgtcaac aagtgcatcg   2580
gcttcagcgt caacgagtgc gtctgagtca gcatcaacga gtgcgtcagc ctcagcaagc   2640
acatcagctt ctgaatctgc atcaaccagt gcgtcagcct cagcatcgac aagcgcctca   2700
gcttcagcaa gtaccagtgc gtcagcctca gcgtcgacaa gtgcgtcggc ttcagcaagt   2760
accagtgcgt cagcctcagc aagtaccagt gcgtcagcct cagcgtcgac aagtgcgtcg   2820
gcctcaacca gtgcatctga atcggcatca accagtgcgt cagcctcagc aagtactagt   2880
gcatcagctt cagcatcaac gagtgcggtt gcttcagcat caaccagtgc atcagagtca   2940
gcaagtacca gtgcgtcagc ttccgcatca acaagtgcct cggcttcagc aagtactagc   3000
gcctcagcct cagcgtcaac aagtgcttca gcttccgcgt caaccagcgc ctcggcctca   3060
gcaagtatct cagcgtctga atcggcatca acaagtgcct cggcttcagc atcaacgagt   3120
gcatcagtct cagcaagcac cagtgcgtcg gcctcagcaa gcaccagcgc gtctgaatcc   3180
gcatcaacca gtgcctcagc ttcagcaagt acctcagcat ctgaatcagc atcaacaagt   3240
gcctcggctt cagcaagcac aagtgcttca gcctcagcaa gtatctcagc gtctgaatcg   3300
gcatcaacga gtgcgtccgc ttcagcaagt actagcgcct cagcatcagc gtcaacaagt   3360
gcttcggctt cagcgtcaac gagtgcgtct gagtcagcat caacgagtgc gtcagcctca   3420
gcaagcacat cagcttctga atctgcatca accagtgcgt cagcctcagc atcgacaagc   3480
gcctcagctt cagcaagtac cagtgcgtca gcctcagcaa gtaccagtgc ttcagcctca   3540
gcgtcgacaa gtgcgtcggc ctcaaccagt gcatctgaat cggcatcaac cagtgcgtca   3600
gcctcagcaa gtactagcgc ctcagcctca gcatcaacga gtgcgtccgc ttcagcaagt   3660
actagtgcat cagcttcagc aagtactagc gcctcagcct cagcgtcgac aagcgcctca   3720
gcttcagcaa gtaccagtgc gtcagcctca gcgtcgacaa gtgcgtcggc ttcagcaagt   3780
acctcagcgt ctgaatcagc atcaacaagt gcgtcggctt cagcatcaac gagtgcatca   3840
gcttcagcat caacaagtgc ttcagcttca gcaagtacca gtgcgtcggc ttcagcatca   3900
acgagtgctt cagtctcagc gtcaaccagt gcctctctga atccgcatca aagtgcctcg   3960
gcttcagcaa gcaccagtgc ttcggcttca gcgtcaacga gtgcgtctga gtcagcatca   4020
acgagtgcgt cagcctcagc aagcacatca gcttctgaat ctgcatcaac cagtgcgtca   4080
gcttccgcat caacaagcgc ctcggcctca gcaagtacaa gtgcttcagc ctcagcatca   4140
accagtgcat cagcttcagc ctcaacaagt gcttcagcct cagcgtcaac cagtgcctcg   4200
gcttcagcaa gtaccagtgc gtcagcttca gcaagcacaa gtgcgtcagc ttcagcatca   4260
accagtgctt cggcttcggc atcaacaagt gcctcagcat cagcatcaac gagtgcgtca   4320
gcctcagcaa gtactagtgc atcagcatca gcatcaacca gtgcatcagc ctcagcaagt   4380
atctcagcgt ctgaatcggc atcaacgagt gcatcagcat cagcatcaac gagtgcatcg   4440
gcttcagcgt caaccagtgc atcagtctca gcaagcacca gtgcgtcagc ttcagcatca   4500
acgagtgcct cagcctcagc aagtatctca gcgtctgaat cggcatcaac gagtgcgtca   4560
gcctcagcaa gtactagtgc atcggcttca gcaagcacca gtgcgtcggc ttcagcatca   4620
accagtgcct cagcctcagc aagtatctca gcgtctgaat cggcatcaac gagtgcgtca   4680
gcctcagcaa gtactagtgc atcagcatca gcatcaacga gtgcatcggc ttcagcaagt   4740
accagcgcct cagcttcagc aagcaccagt gcgtcagcct cagcaagtac cagcgcctca   4800
gcctcagcaa gcaccagtgc ctcagcttca gcaagtacca gtgcgtcagc ctcagcgtcg   4860
acaagtgcgt cggcttcagc aagtacctca gcgtctgaat cagcatcaac gagtgcatca   4920
```

```
gcttcagcat caacaagtgc ttcagcttca gcaagtacca gtgcgtcggc ttcagcatca   4980
acgagtgctt cagtctcagc gtcaaccagt gcctctgaat cagcatcaac aagtgcctcg   5040
gcttcagcaa gcaccagtgc gtcggcttca gcaagtacta gtgcatcggc ttcagcatcg   5100
acaagtgcgt ctgaatcggc atcaacgagt gcttcggctt cagcatcaac gagtgcgtca   5160
gcctcagcaa gcacatcagc ttctgaatct gcatcaacca gtgcgtccgc ttcagcgtca   5220
accagtgcgt cggcttcagc gtcgacaagt gcttcggctt cagcatcaac gagtgcgtcg   5280
gcctcagcaa gcgcaagtac ctcagcgtca gcttccgcct caaccagtgc gtcggcttca   5340
gcaagcacaa gtgcgtcagc ctcagcaagt atctcagcgt ctgaatcggc atcaacgagt   5400
gcgtctgagt cagcatcaac gagtacgtca gcctcagcaa gcacatcagc ttctgaatct   5460
gcatcaacca gtgcgtcagc ctcagcatca caagcgcct cagcttcagc aagtaccagt   5520
gcttcagcct cagcgtcgac aagtgcgtcg gcctcaacca gtgcatctga atcggcatca   5580
accagtgcgt cagcctcagc aagtactagt gcatcagctt cagcatcaac gagtgcatcg   5640
gcttcagcat caaccagtgc ctcggcttca gcgtcaacca gtgcgtcagc ttcagcaagt   5700
accagtgctt cagtctcagc atcaacaagt gcttcagcct cagcatcgac aagtgcctcg   5760
gcttcagcaa gcacatcagc atctgaatca gcgtcgacaa gcgcctcagc ttcagcaagt   5820
accagtgcgt cagcctcagc gtcgacaagt gcgtcagcct cagcaagtac tagtgcatca   5880
gcttcagcat caacgagtgc atcggcttcg gcgtcaacca gtgcatcaga gtcagcaagt   5940
accagtgcgt cagcttccgc atcaacaagt gcctccggct tcagcaagca cagtgcgtcg   6000
gcttcagcaa gtactagcgc ctcagcctca gcctcaacca gtgcgtcagc ctcagcaagt   6060
atctcagcgt ctgaatcggc atcaacgagt gcgtccgctt cagcaagtac tagcgcctca   6120
gcctcagcgt caacaagtgc atcggcttca gcgtcaacga gtgcgtctga atcggcatca   6180
acgagtgcgt ccgcttcagc aagtactagc gcctcagcct cagcgtcaac aagtgcatcg   6240
gcttcagcat caacgagtgc gtccgcttca gcaagtacta gcgcctcagc ctcagcgtca   6300
acaagtgcat cggcttcagc gtcaacgagt gcgtctgagt cagcatcaac gagtgcgtca   6360
gcctcagcaa gcacatcagc ttctgaatct gcatcaacca gtgcgtcagc ctcagcatcg   6420
acaagcgcct cagcttcagc aagtaccagt gcgtcagcct caccagtgca agtgcaagtg   6480
gcttcagcaa gtaccagtgc gtcagcctca gcaagtacca gtgcgtcagc ctcagcgtcg   6540
acaagtgcgt cggcctcaac cagtgcatct gaatcggcat caaccagtgc gtcagcctca   6600
gcaagtacta gtgcatcagc ttcagcatca acgagtgcat cggcttcagc atcaaccagt   6660
gcatcagagt cagcaagtac cagtgcgtca gcttccgcat caacaagtgc ctcggcttca   6720
gcaagtacta gcgcctcagc ctcagcgtca acaagtgctt cagcttccgc gtcaaccagc   6780
gcctcggcct cagcaagtat ctcagcgtct gaatcggcat caacaagtgc ctcggcttca   6840
gcatcaacga gtgcatcagt ctcagcaagc accagtgcgt cggcctcagc aagcaccagc   6900
gcgtctgaat ccgcatcaac cagtgcctca gcttcagcaa gtacctcagc atctgaatca   6960
gcatcaacga gtgcatcggc ttcagcaagc acaagtgctt cagcctcagc aagtatctca   7020
gcgtctgaat cggcatcaac gagtgcgtcc gcttcagcaa gtactagcgc ctcagcatca   7080
gcgtcaacaa gtgcttcggc ttcagcgtca acgagtgcgt ctgagtcagc atcaacgagt   7140
acgtcagcct cagcaagcac atcagcttct gaatctgcat caaccagtgc gtcagcctca   7200
gcatcgacaa gcgcctcagc ttcagcaagt accagtgcgt cagcctcagc aagtaccagt   7260
gcttcagcct cagcgtcgac aagtgcgtcg gcctcaacca gtgcatctga atcggcatca   7320
accagtgcgt cagcctcagc aagtactagc gcctcagcct cagcatcaac gagtgcgtcc   7380
gcttcagcaa gtactagtgc atcagcatca gcatcaacga gtgcatcggc ttcagcaagt   7440
accagcgcct cagcttcagc aagcaccagt gcgtcagcct cagcaagtac cagcgcctca   7500
gcctcagcaa gcaccagtgc ctcagcgttc gcaagtacca gtgcgtcagc ctcagcgtcg   7560
acaagtgcgt cggcttcagc aagtacctca gcgtctgaat cagcatcaac gagtgcatca   7620
gcttcagcat caacaagtgc ttcagcttca gcaagtacca gtgcgtcggc ttcagcatca   7680
acgagtgctt cagtctcagc gtcaaccagt gcctctgaat cagcatcaac aagtgcctcg   7740
gcttcagcaa gcaccagtgc gtcggcttca gcaagtacta gtgcatcggc ttcagcatcg   7800
acaagtgcgt ctgaatcggc atcaacgagt gcttcggctt cagcatcaac gagtgcgtca   7860
gcctcagcaa gcacatcagc ttctgaatct gcatcaacca gtgcgtccgc ttcagcgtca   7920
accagtgcgt cggcttcagc gtcgacaagt gcttcggctt cagcatcaac gagtgcgtcg   7980
gcctcagcaa gcgcaagtac ctcagcgtca gcttccgcct caaccagtgc gtccgcttca   8040
gcaagcacaa gtgcgtcagc ctcagcaagt atctcagcgt ctgaatcggc atcaacgagt   8100
gcgtcggcct cagcaagcgc aagtacctca gcgtcagctt ccgcctcaac cagtgcgtcg   8160
gcttcagcaa gcacaagtgc gtcagcctca gcaagtatct cagcgtctga atcggcatca   8220
acgagtgcgt ctgagtcagc atcaacgagt acgtcagcct cagcaagcac atcagcttct   8280
gaatcggcat caaccagtgc gtcagcctca gcatcgacaa gcgcctcagc ttcagcaagt   8340
accagtgctt cagcctcagc gtcgacaagt gcgtcggcct caaccagtgc atctgaatcg   8400
gcatcaacca gtgcgtcagc ctcagcaagt actagtgcat cagcttcagc atcaacgagt   8460
gcatcggctt cagcatcaac cagtgcctcg gcttcagcgt caaccagtgc gtcagcttca   8520
gcaagtacca gtgcttcagt ctcagcatca acaagtgctt cagcctcagc atcgacaagt   8580
gcctcggctt cagcaagcgt gaatcagcgt cgacaagtgc gtcggcctca   8640
accagtgcat ctgaatcggc atcaaccagt gcgtcagcct cagcaagtac tagtgcatca   8700
gcttcagcat caacgagtgc atcggcttcg gcgtcaacca gtgcatcaga gtcagcaagt   8760
accagtgcgt cagcttccgc atcaacaagt gcctcggctt cagcaagcac atcagcatct   8820
gaatcagcgt caaccagtgc ttcggcttca gcaagtacca gtgcttcagc ttcagcatca   8880
accagcgcct cggcctcagc aagcacctca gcttctgaat cggcctcaac cagcgcctcg   8940
gcctcagcaa gcacctcagc ttctgaatcg gcctcaacca gcgcctcagc ctcagcatca   9000
```

```
acgagtgctt cggcttcagc aagcacaagc gcctcgggtt cagcatcaac gagtacgtca   9060
gcttcagcgt caaccagtgc ttcagcctca gcatcaacaa gtgcgtcagc ctcagcaagt   9120
atctcagcgt ctgaatcggc atcaacgagt gcgtctgagt cagcatcaac gagtacgtca   9180
gcctcagcaa gcacctcagc ttctgaatcg gcctcaacca gtgcgtcagc ctcagcatcg   9240
acaagcgcct cagcttcagc aagtaccagt gcttcagcct cagcgtcgac aagtgcgtcg   9300
gcctcaacca gtgcatctga atcggcatca accagtgcgt cagcctcagc aagtactagt   9360
gcatcggctt cagcatcaac cagtgcctcg gcttcagcgt caaccagtgc gtcagcttca   9420
gcaagtacca gtgcttcagt ctcagcatca acaagtgctt cagcctcagc atcgacaagt   9480
gcctcggctt cagcaagcac atcagcatct gacagcgt cgacaagcgc ctcagcttca   9540
gcaagtacca gtgcgtcagc ctcagcgtcg acaagtgcgt cagcctcagc aagtactagt   9600
gcatcagctt cagcatcaac gagtgcatcg gcttcggcgt caaccagtgc atcagagtca   9660
gcaagtacca gtgcgtcagc ttccgcatca acaagtgcct cggcttcagc aagcaccagt   9720
gcgtcggctt cagcaagtac tagcgcctca gcctcagcct caaccagtgc gtcagcctca   9780
gcaagtatct cagcgtctga atcggcatca acgagtgcgt ccgcttcagc aagtactagc   9840
gcctcagcct cagcgtcaac aagtgcatcg gcttcagcgt caacgagtgc gtctgaatcg   9900
gcatcaacga gtgcgtccgc ttcagcaagt actagcgcct cagcctcagc gtcaacaagt   9960
gcatcggctt cagcatcaac gagtgcgtcc gcttcagcaa gtactagcgc ctcagcctca   10020
gcgtcaacaa gtgcatcggc ttcagcgtca acgagtgcgt ctgagtcagc atcaacgagt   10080
gcgtcagcct cagcaagcac atcagcttct gaatctgcat caaccagtgc gtcagcctca   10140
gcatcgacaa gcgcctcagc ttcagcaagt accagtgcgt cagcctcagc gtcgacaagt   10200
gcgtcggctt cagcaagtac cagtgcgtca gcctcagcaa gtaccagtgc gtcagcctca   10260
gcgtcgacaa gtgcgtcggc ctcaaccagt gcatctgaat cggcatcaac cagtgcgtca   10320
gcctcagcaa gtactagtgc atcagcttca gcatcaacga gtgcatcggc ttcagcatca   10380
accagtgcat cagagtcagc aagtaccagt gcgtcagctt ccgcatcaac aagtgcctcg   10440
gcttcagcaa gtactagcgc ctcagcctca gcgtcaacaa gtgcttcagc ttccgcgtca   10500
accagcgcct cggcctcagc aagtatctca gcgtctgaat cggcatcaac aagtgcctcg   10560
gcttcagcat cagtcagcgt catcagca gcgtcgcggc gtcagcaagc   10620
accagcgcgt ctgaatccgc atcaaccagt gcctcagctt cagcaagtac ctcagcatct   10680
gaatcagcat caacaagtgc ctcggcttca gcaagcacaa gtgcttcagc ctcagcaagt   10740
atctcagcgt ctgaatcggc atcaacgagt gcgtccgctt cagcaagtac tagcgcctca   10800
gcatcagcgt caacaagtgc ttcggcttca gcgtcaacga gtgcgtctga gtcagcatca   10860
acgagtacgt cagcctcagc aagcacatca gcttctgaat ctgcatcaac cagtgcgtca   10920
gcctcagcat cgacaagcgc ctcagcttca gcaagtacca gtgcgtcagc ctcagcaagt   10980
accagtgctt cagcctcagc gtcgacaagt gcgtcggcct caaccagtgc atctgaatcg   11040
gcatcaacca gtgcgtcagc ctcagcaagt actagcgcct cagcctcagc atcaacgagt   11100
gcgtccgctt cagcaagtac tagtgcatca gcttcagcaa gtactagcgc ctcagcctca   11160
gcgtcgacaa gcgcctcagc ttcagcaagt accagtgcgt cagcctcagc gtcgacaagt   11220
gcgtcggctt cagcaagtac ctcagcgtct gaatcagcat caacaagtgc gtcggcttca   11280
gcatcaacga gtgcatcagc ttcagcatca acaagtgctt cagcttcagc aagtaccagt   11340
gcgtcggctt cagcatcaac gagtgcttca gtctcagcgt caaccagtgc ctctgaatcc   11400
gcatcaacaa gtgcctcggc ttcagcaagc accagtgctt cggcttcagc gtcaacgagt   11460
gcgtctgagt cagcatcaac gagtgcgtca gcctcagcaa gcacatcagc ttctgaatct   11520
gcatcaacca gtgcgtcagc ttccgcatca acaagcgcct cggcctcagc aagtacaagt   11580
gcttcagcct cagcatcaac cagtgcatca gcttcagcct caaccagtgc ttcagcctca   11640
gcgtcaacca gtgcgtcggc ttcagcaagt accagtgcgt cagcttcagc aagcacaagt   11700
gcgtcagctt cagcatcaac cagtgcttcg gcttcggcat caacaagtgc ctcagcatca   11760
gcatcaacga gtgcgtcagc ctcagcaagt actagtgcat cagcatcagc atcaaccagt   11820
gcatcagcct cagcaagtat ctcagcgtct gaatcggcat caacgagtgc atcagcatca   11880
gcatcaacga gtgcatcggc ttcagcgtca accagtgcat cagtctcagc aagcaccagt   11940
gcgtcggctt cagcatcaac gagtgcctca gcctcagcaa gtatctcagc gtctgaatcg   12000
gcatcaacga gtgcgtcagc ctcagcaagt actagtgcat cggcttcagc aagcaccagt   12060
gcgtcggctt cagcatcaac cagtgcctca gcctcagcaa gtatctcagc gtctgaatcg   12120
gcatcaacga gtgcgtcagc ctcagcaagt actagtgcat cagcatcagc atcaacgagt   12180
gcatcggctt cagcaagtac cagcgcctca gcttcagcaa gcaccagtgc gtcagcctca   12240
gcaagtacca gcgcctcagc ctcagcaagc accagtgcct cagcttcagc aagtaccagt   12300
gcgtcagcct cagcgtcgac aagtgcgtcg gcttcagcaa gtacctcagc gtctgaatca   12360
gcatcaacga gtgcatcagc ttcagcatca acaagtgctt cagcttcagc aagtatctca   12420
gcgtctgaat cggcatcaac gagtgcgtcc gcttcagcaa gtactagcgc ctcagcatca   12480
gcgtcaacaa gtgcttcggc ttcagcgtca acgagtgcgt ctgagtcagc atcaacgagt   12540
acgtcagcct cagcaagcac atcagcttct gaatctgcat caaccagtgc gtcagcctca   12600
gcatcgacaa gcgcctcagc ttcagcaagt accagtgcgt cagcctcagc aagtaccagt   12660
gcttcagcct cagcgtcgac aagtgcgtcg gcctcaacca gtgcatctga atcggcatca   12720
accagtgcgt cagcctcagc aagtactagc gcctcagcaa cagtgcgtcc gagtgcgtcc   12780
gcttcagcaa gtactagtgc atcagcttca gcaagtacta gcgcctcagc ctcagcgtcg   12840
acaagcgcct cagcttcagc aagtaccagt gcgtcagcct cagcgtcgac aagtgcgtcg   12900
gcttcagcaa gtacctcagc gtctgaatca gcatcaacaa gtgcgtcggc ttcagcatca   12960
acgagtgcat cagcttcagc atcaacaagt gcttcagctt cagcaagtac cagtgcgtcg   13020
gcttcagcat caacgagtgc ttcagtctca gcgtcaacca gtgcctctga atccgcatca   13080
```

```
acaagtgcct cggcttcagc aagcaccagt gcttcggctt cagcgtcaac gagtgcgtct    13140
gagtcagcat caacgagtgc gtcagcctca gcaagcacat cagcttctga atctgcatca    13200
accagtgcgt cagcttccgc atcaacaagc gcctcggcct cagcaagtac aagtgcttca    13260
gcctcagcat caaccagtgc atcagcttca gcctcaacaa gtgcttcagc ctcagcgtca    13320
accagtgcct cggcttcagc aagtaccagt gcgtcagctt cagcaagcac aagtgcgtca    13380
gcttcagcat caaccagtgc ttcggcttcg gcatcaacaa gtgcctcagc atcagcatca    13440
acgagtgcgt cagcctcagc aagtactagt gcatcagcat cagcatcaac cagtgcatca    13500
gcctcagcaa gtatctcagc gtctgaatcg gcatcaacga gtgcatcagc atcagcatca    13560
acgagtgcat cggcttcagc gtcaaccagt gcatcagtct cagcaagcac cagtgcgtcg    13620
gcttcagcat caacgagtgc ctcagcctca gcaagtatct cagcgtctga atcggcatca    13680
acgagtgcgt cagcctcagc aagtactagt gcatcggctt cagcaagcac cagtgcgtcg    13740
gcttcagcat caaccagtgc ctcagcctca gcaagtatct cagcgtctga atcggcatca    13800
acgagtgcgt cagcctcagc aagtactagt gcatcagcct cagcatcaac gagtgcatcg    13860
gcttcagcaa gtaccagcgc ctcagcttca gcaagcacca gtgcgtcagc ctcagcaagt    13920
accagcgcct cagcctcagc aagcaccagt gcctcagctt cagcaagtac cagtgcgtca    13980
gcctcagcat caacaagtgc ttcagcttcg gcctcaacaa gtgcgtcagc ttcagcatca    14040
acgagtgcgt cggcttcagc aagcaggcct cagcaagcac cagcaagcac cagtgcttca    14100
gcttcagcat caacaagtgc gtcagcttca gcaagtacat cagtttcaaa ttcagcaaac    14160
cattcgaact cacaagttgg aaatacttct ggatcgacag gtaaatccca aaaagaattg    14220
cctaatacag gtactgagtc gtcaattgga tctgtgttac ttggagttct agcagctgtt    14280
acaggtattg gattggttgc gaaacgccgt aaacgtgatg aagaagag              14328


SeqID 74
atgtcaaacg aaaaaaacac aaacactaac gtagaaaaga aagatgctac tgttgtagct     60
cacgaaatca aaggggaact tacttacgaa gataaagtta tccaaaaaat cattggtctt    120
tcactagaaa acgtttcagg tcttttggga atcgatggtg gtttcttctc aaatcttaaa    180
gaaaaaatcg ttaacagcga tgacgtaaca agtgtgtta acgtagaagt tggtaaaaca    240
caagttgcag ttgacttaaa cgttattgtt gagtaccaaa aaaatgttcc agctttatat    300
tcagaaatca gagaaatcgt atcttcagaa gttgctaaaa tgactgactt ggaaattgtt    360
gaaatcaacg taaacgttgt cgacatcaaa actaaagaac agcatgaagc agactcagta    420
agccttcaag atcgcgtatc tgacgttgct gaatcaacag gagaattcac ttcagaacaa    480
ttcgaaaaag ctaaatctgg tcttggatct ggtttctcaa ctgttcaaga aaaagttagc    540
gaaggtgtag aagctgttaa aggtgcagca aatggtgtag tatctcacga aaacactcgt    600
gtaaac                                                              606


SeqID 75
atgacaaaag aaaaaaatgt aattttgact gctcgcgata ttgtcgtgga atttgacgtt     60
cgtgacaaag tattgacagc cattcgcggc gtttcccttg aactagtcga aggagaagta    120
ttagccttgg taggtgagtc aggatcaggt aaatctgttt tgacaaagac cttcacaggt    180
atgctcgaag aaaatggtcg tattgcccaa ggtagtattg actaccgtgg tcaggacttg    240
acagctttat cttctcacaa ggattgggaa caaattcgtg gtgctaagat tgcgactatc    300
ttccaggacc caatgactag tttggacccc attaaaacaa ttggtagtca gattacagaa    360
gttattgtaa aacaccaagg aaaaacagct aaagaagcga aagaattggc cattgactac    420
atgaataagg ttggcattcc agacgcagat agacgtttta atgataaccc attccaatat    480
tctggaggaa tgcgtcaacg tatcgttatt gctattgccc ttgcctgccg acctgatgtc    540
ttgatctgtg atgagccaac aactgccttg gatgtaacta ttcaagctca gattattgat    600
ttgctaaaat ctttacaaaa cgagtatcat ttcacaacaa tctttattac ccacgacctt    660
ggtgtggtgg caagtattgc ggataaggta gcggttatgt atgcaggaga aatcgttgag    720
tatggaacgg ttgaggaagt cttctatgac cctcgccatc catatacatg gagtctcttg    780
tctagcttgc ctcagcttgc tgatgataaa ggggatcttt actcaatccc aggaacacct    840
ccgtcacttt atactgacct gaaaggggat gcttttgcct tgcgttctga ctacgcaatg    900
cagattgact tcgaacaaaa agctcctcaa ttctcagtat cagagacaca ttgggctaaa    960
acttggcttc ttcatgagga tgctccgaaa gtagaaaaac cagctgtgat tgcaaatctc    1020
catgataaga tccgtgaaaa aatgggattt gcccatctgg ctgac                  1065


SeqID 76
atgaaaaaaa atcgtgtatt tgctacagca ggtcttgttt tattagcagc aggtgtactt     60
gcagcatgca gttcttcaaa atcatctgat tcatcagccc ctaaagctta tggctatgtt    120
tatacagcag acccagaaac cttggactac ctgatttcaa gtaaaaatag tacaacagta    180
gtgacttcaa atgggattga tggtttattc actaacgata attacggtaa tcttgctcct    240
gcagttgcag aggattggga agtctctaag gatggtttga cctacactta taagattcgt    300
aaaggggtta aatggttac ctctgatgga gaagaatatg cagaggtgac ggctaaagat    360
ttcgtgaacg gtttaaaaca cgcagcagat aaaaaatcag aagctatgta tttagctgaa    420
aattcggtta aaggcttggc agattatcta tcaggaactt caacagattt ttcaacagtt    480
ggtgtcaagg cggttgatga ttatacgtta caatacactt gaaccagcc tgaaccgttc    540
tggaactcta agttgaccta ttctatttc tggcctctga atgaagaatt cgaaacatca    600
aaaggaagcg attttgctaa accaacagat ccgacatcct tgctttataa tggtccattc    660
ttgttgaaag ggttgactgc aaaatcttct gtagagtttg taaaaaatga gcaatattgg    720
```

```
gataaagaaa atgtccacct agatactatc aatctagctt actatgatgg atcagatcag     780
gagtcgctag agcgtaactt cactagtgga gcttatagtt atgcccgtct ttaccctacc     840
agctccaact attctaaggt tgcagaagaa tacaaggaca atatctatta cacacaatca     900
ggctctggga ttgctggtct gggtgtgaat attgatcgcc aaagttacaa ctatacttct     960
aaaactacag attcagagaa agtagctact aagaaggcat tgcttaacaa agatttccgt    1020
caagccttga attttgctct tgatcgctca gcttactcag ctcaaatcaa tggtaaagat    1080
ggagcagctt tagcagttcg taatttattt gtaaaaccag actttgtttc agctggtgag    1140
aagacctttg gtgatttagt cgctgctcaa cttcctgctt atggtgatga gtggaaaggt    1200
gtgaatttag ctgatgggca ggatggttta ttcaatgctg acaaggccaa ggcagagttt    1260
gcgaaagcta agaaagcttt agaagcagac ggcgttcagt ttcctattca tctggacgtt    1320
ccagtagacc aagcatcaaa aaactacata tctcgtattc agtcctttaa acaatctgta    1380
gaaacagttc ttggtgttga aaatgtcgtt gttgatattc aacaaatgac aagtgatgaa    1440
ttccttaata ttacttacta tgctgccaat gcttcatctg aggattggga tgtatcagga    1500
ggagtttcat gggggccaga ctatcaagac ccatctactt acctggatat tttaaaaaca    1560
actagcagtg aaactacaaa aacatattta ggatttgata atccaaatag cccttcagta    1620
gttcaagttg gtttgaaaga atacgataaa ttagttgatg aagctgccag agagacaagc    1680
gacttgaatg tccgttatga aaaatatgca gcggctcaag catggttgac agatagttca    1740
ctctttattc ctgctatggc ttcttctggt gcagcaccag tgctttcacg aattgttcca    1800
tttactggag cttctgcgca aacaggctct aaggggtcag atgtttactt caaatatttg    1860
aaatcacaag ataaagtggt gactaaggaa gagtatgaaa aagctcgtga aaaatggttg    1920
aaagaaaaag ctgaatcaaa tgagaaagct caaaaagaat tggcaagtca tgtgaag      1977
```

SeqID 77

```
atggaaatta atgtgagtaa attaagaaca gatttgcctc aagtcggcgt gcaaccatat      60
aggcaagtac acgcacactc aactgggaat ccgcattcaa ccgtacagaa tgaagcggat     120
tatcactggc ggaaagaccc agaattaggt tttttctcgc acattgttgg gaacggttgc     180
atcatgcagg taggacctgt tgataatggt gcctgggacg ttgggggcgg ttggaatgct     240
gagacctatg cagcgggttga actgattgaa agccattcaa ccaaagaaga gttcatgacg     300
gactaccgcc tttatatcga actcttacgc aatctagcag atgaagcagg tttgccgaaa     360
acgcttgata cagggagttt agctggaatt aaaacgcacg agtattgcac gaataaccaa     420
ccaaacaacc actcagacca cgttgaccct tatccatatc ttgctaaatg gggcattagc     480
cgtgagcagt ttaagcatga tattgagaac ggcttgacga ttgaaacagg ctggcagaag     540
aatgacactg gctactggta cgtacattca gacggctctt atccaaaaga caagtttgag     600
aaaatcaatg gcacttggta ctactttgac agttcaggct atatgcttgc agaccgctgg     660
aggaagcaca cagacggcaa ctggtactgg ttcgacaact caggcgaaat ggctacaggc     720
tggaagaaaa tcgctgataa gtggtactat ttcaacgaag aaggtgccat gaagacaggc     780
tgggtcaagt acaaggacac ttggtactac ttagacgcta aagaaggcgc catggtatca     840
aatgccttta tccagtcagc ggacggaaca ggctggtact acctcaaacc agacggaaca     900
ctggcagaca agccagaatt cacagtagag ccagatggct tgattacagt aaaa          954
```

SeqID 78

```
atgaaaaaaa aatattggac tttagcgata ttattctttt gtttgttcaa taattctgtt      60
actgctcaag aaataaacctaa aaatcttgat ggcaatataa ctcacactca gactagcgaa     120
agttttttctg aatctgatga aaaacaggtt gactattcta ataaaaatca agaagaagta     180
gaccaaaata aatttcgtat tcaaatcgat aagacagaat tatttgtaac aacagataaa     240
catttagaaa aaaactgttg taaattggaa cttgaaccac aaataaataa cgatattgtt     300
aactctgaaa gtaataattt actaggcgaa gataatttag ataataaat taaggaaaat     360
gtttctcatc tagataatag aggaggaaat atagagcatg acaaagataa cttagaatcg     420
tcgattgtaa gaaaatatga atgggatata gataaagtta ctggtggagg cgaaagttat     480
aaattatatt ctaaaagtaa ttctaaagtt tcaattgcta ttttagattc aggagtcgat     540
ttacaaaata ctggattact gaaaaatctt tcaaatcact caaaaaacta tgtccccaat     600
aaaggatatt taggaaaaga ggagggagag gaaggaataa tatcagatat tcaagataga     660
ttaggtcatg gtacggctgt tgtagctcaa attgtagggg atgacaatat taatggagta     720
aatcctcacg ttaatattaa cgtctcataga atatttggta agtcgtcagc tagtccagat     780
tggattgtaa aagcaatttt tgatgctgta gatgatggca atgatattat caatcttagt     840
actggacaat atttaatgat tgatggagaa tatgaggacg gaacaaatga ttttgaaaca     900
ttttttgaagt ataaaaaggc tattgattac gcgaatcaaa aaggagtaat tatagtagct     960
gcattaggga atgactccct aaatgtatca aatcagtcag atttattgaa acttattagt    1020
tcacgcaaaa aagtaagaaa accaggatta gtagttgatg ttccaagtta tttctcatct    1080
acaatttcgg tcggaggcat agatcgctta ggtaatttat cagattttag caataaaggg    1140
gattctgatg caatatatgc gcctgcaggc tcaacattat ctctttcaga attaggactt    1200
aataacttta ttaatgcaga aaaatataaa gaagattgga ttttttcggc aacactagga    1260
ggatatacgt atctttatgg aaactcattt gctgctccta agtttctgg tgcgattgca    1320
atgattattg ataaatacaa attaaaagat cagccctata attatatgtt tgtaaaaaaa    1380
ttctggaaga aacattacca g                                              1401
```

SeqID 79

```
atgaaaaaag atgagttatt tgaaggcttt tacctaatca aatcagctga cctgaggcaa      60
```

```
actcgagctg ggaaaaacta cctagccttt accttccaag atgatagtgg cgagattgat   120
gggaagctct gggatgccca acctcataac attgaggcct ttaccgcagg taaggttgtc   180
cacatgaaag gacgccgaga agtttataac aataccctc aagtcaatca aattactctc    240
cgcctgcctc aagctggtga acccaatgac ccagctgatt tcaaggtcaa gtcaccagtt   300
gatgtcaagg aaattcgtga ctacatgtcg caaatgattt tcaaaattga aaatcctgtc   360
tggcaacgga ttgtccgaaa tctctacacc aagtatgata aggaattcta ctcctatcca   420
gctgccaaga ccaaccacca tgcctttgaa acgggcttgg cctatcatac ggcgaccatg   480
gtgcgtttgg cagacgctat tagcgaagtt tatcctcagc tcaataagag cctgctctat   540
gcggggatta tgttgcatga cttagctaag gtcatcgagt tgacggggcc agaccagaca   600
gagtacacag tgcgaggtaa tcttcttgga catatcgctc tcattgatag cgaaattacc   660
aagacagtta tggaactcgg catcgatgat accaaggaag aagtcgtttt gcttcgtcat   720
gtcatcctca gtcaccacgg cttgcttgag tatggaagcc cagtccgtcc acgcattatg   780
gaagcagaga ttatccatat gattgacaat ctggatgcaa gcatgatgat gatgtcaaca   840
gctcttgctt tggtggataa aggagagatg accaataaaa tcttcgctat ggataatcgt   900
tccttctata aaccagattt agat                                         924
```

SeqID 80
```
gtgacgattc taggaaaaga tacagttcaa caatctgcga aaggtgaatc tgtaactcaa    60
gaagctacac cagagtataa gctagaaaat acaccaggtg gagataaggg aggcaatact   120
ggaagctcag atgctaatgc gaatgaaggc ggtggtagcc aggcgggtgg atcagctcac   180
acaggttcac aaaactcagc tcaatcacaa gcttctaagc aattagctac tgaaaaagaa   240
tcagctaaaa atgccattga aaaagcagcc aaggacaagc aggatgaaat caaaggcgca   300
ccgctttctg ataaagaaaa agcagaactt ttagcaagag tggaagcaga aaaacaagca   360
gctctcaaag agattgaaaa tgcgaaaact atggaagatg tgaaggaagc agaaacgatt   420
ggagtgcaag ccattgccat ggttacagtt cctaagagac cagtggctcc taatgctgct   480
cctaagacaa caagtgcacc gcaagcaact gcaggaacaa tgcaagatgt tacctaccag   540
tcacctgctg gcaaacaatt acctaacaca ggttcagcat caagtgcagc acttgctagt   600
cttggtctag tggtggcaac aagtggtttt gctttgctag gaagaaagac tagacgtaga   660
aaa                                                               663
```

SeqID 81
```
atgaatgcag atgatacagt aaccatttat gatgtcgctc gtgaagcagg tgtttccatg    60
gcgacggtca gccgtgtggt caatggcaat aaaaatgtaa aagagaatac ccgtaaaaaa   120
gtgctagagg taattgatcg tttggattat cgtccaaatg cagttgcgcg tggtcttgca   180
agtaaaaaga caaccactgt cggtgtcgtg attccaaata ttaccaatgg ttattttttca  240
agtttggcta aggggattga tgatattgca gaaatgtaca agtacaatat tgtcctagct   300
aatagcgatg aagataacga gaaagaagtt tctgttgtca ataccctctt ttcaaagcag    360
gtagatggca ttatctatat ggggtatcac ttgacagata aaattcgctc agaattttcg   420
cgttcacgta ctccgattgt tctcgcagga actgtcgatg ttgagcacca gttgccaagt   480
gtcaatattg actataagca agcaacaatt gatgcagtga gttaccttgc taaagaaaat   540
gagcgtattg ctttcgttag cggtccgcta gtggatgaca tcaatggtaa ggttcgttta   600
gttggctaca aggaaacctt gaaaaaagca ggaatcactt atagtgaggg tttggtattt   660
gaatctaaat atagctatga tgatggttac gccttagcag acgtttgat ttcatcaaat    720
gcaactgcag cagttgtgac aggtgatgag ttggcagcag gagtcttgaa cggttttggct  780
gataaggggt tttctgtgcc agaagatttt gaaattatta ctagtgatga ttcacaaatc   840
tcacgcttta cccgtccaaa cttgacaacg attgcccaac ctctttatga ccttggtgcc   900
attagtatgc gtatgttgac caagattatg cataaggaag agttggaaga acgtgaagtt   960
ctcttacctc atggtttgac agaacgtagc tcaacacgaa aacgtaaa             1008
```

SeqID 82
```
atgaaaaaaa agttagtatt tcctaatctg ttttggtggg gagctgcttc tagcggacct    60
cagacagaag gtcaatatgg aaaagtacat gaaaatgtga tggactactg gttcaaaacg   120
catccagaga atttttttcga taatgtcgga cctcttgtag ccagtaactt ttttcatact   180
tacaccgaag atttccactt gatgaaggaa attggagtta attctttccg cacttccatc   240
caatggagtc gactcatcaa gaatttagag acaggtgagc ctgatccaaa aggtattgct   300
ttctacaatg ccatcattga agaagctaaa aagaaccaga tggatcttgt gatgaattta   360
catcattttg atttaccagt ggaacttctt caaaaatacg gtggttggga aagcaaacat   420
gtagtggagt tattcgtgaa gtttgccaag actgctttca catgctttgg agataaggtt   480
cattactgga caactttcaa tgagccaatg gtcattccag aagcaggata cttatatgct   540
ttccattatc caaatctaaa aggaaaggga aaagaggccg tacaagtcat ctataatcta   600
aaccttgcta gtgcaaaagt gattcaacta tatcgctcat tagaacttga tggaaagatt   660
gggattattt aaaactgac acctgcttat ccaagaagta attctccaga agacttagaa   720
gcaagtcgat ttacagatga cttctttaac aaagtcttct tgaatccagc tgttaaagga   780
actttcccag aaagattggt aaaacagcta gagagagatg gcgtgttatg gagtcatacc   840
gaaaaagagc ttcaactgat gaaatcaaat acggttgatt ttcttggagt aaactactac   900
catccaaaac gtgttcaagc acaagcaaat cctgaggaat atcagacgcc ctggatgcca   960
gaccaatact tcaaagagta tgaatggctg gagcgtcgca tgaatccata tcgtggttgg  1020
gaaattttc cgaaagccat ttatgatatt gctatgattg tgaaggaaga atatggtaat  1080
```

```
atcccatggt ttatcagtga aaacggaatg ggtgttgaaa acgaagcacg gtttatcgat    1140
gaaaatggag ttatcgatga cgtgtatcgt attgaatttt atgaagaaca tttaagatgg    1200
ctacataaag ccattgaaga gggaagtcac tgttttggat accacgcttg gaccgcattt    1260
gattgctggt cttggaataa tgcatataag aatcgttacg gatttatctc cgttgattta    1320
gaaacgcaaa agagaaccat caagagctca ggaagatggt atcgcaaagt aagtgacaat    1380
aacggttttg aagtagaaat tgaggag                                        1407
```

SeqID 83
```
gtggaaaatc ttacgaattt ttacgaaaag tatcgtgtct atctgactcg tccacgttta      60
gagcttttgg cagtagttac cattgttttc tgtgctgtac tcgtcttttt tctaaatatt     120
ccaggaaaag gtgtcttaaa actcgataat ggaacgattg tttatgatgg cagtcttgtc     180
cgcggtaaaa tgaatggcca aggtaccatt accttccaaa atggagacca atatacaggt     240
ggcttcaaca atggagcctt caacggaaaa ggtacctttc aatctaaaga aggctggacc     300
tacgaaggtg attttgtaaa tggtcaggct gaaggaaaag ggaaactaac aacagaacaa     360
gaagtcgttt atgaaggaac ttttaaacaa ggcgtttttc aacaaaaa                  408
```

SeqID 84
```
atgttgaata agataagaga ctatttagac tttgctggtt tgcagtaccg taatcctgat      60
aaagcgggag cagagcgaga gaagatgctg gcattccgcc acaaaggaca agaggcccga     120
aaggttttta cagaactggc caaagccttt caagcaagcc atccagaatg gcaactccaa     180
cagactagcc agtggatgaa tcaggcccag cgtttgagac cacattttg ggttatcta      240
cagagagacg gacaagtgac agaacctatg atggccttac gtttgtatgg acatctact      300
gactttggaa tttctttgga agtcagtttc atcgaacgta agaaggatga gcaaacactg     360
ggcaagcagg ccaaagtttt agacattcca accgttaaag ggatttatta tctaacctac     420
tctaatggtc aaagtcaacg gtgggaggcg aatgaagaaa agcgtcgtac tttacgcgag     480
aaggtgagaa gtcaagaagt tcgaaaagtt ttagtgaagg tagatgttcc tatgacagaa     540
aattcgtctg aagaagaaat cgtagaaggc ttattgaagt cttattctaa aattcttccc     600
tattatctag ctacgagaaa a                                              621
```

SeqID 85
```
atggttcaga acagttgttg gcaatcaaag agccataagg tcaaggcttt taccttgtta      60
gaatccctgc ttgccctcat tgtcatcagt gggggattac tccttttca agctatgagt     120
cagctcctca tttcagaagt tcgctaccag caacaaagcg agcaaaagga gtggctcttg     180
tttgtggacc aacttgaggt agaattagac cgttcgcagt tcgaaaaagt agaaggcaat     240
cgcctataca tgaagcaaga tggcaaggac atcgccatcg gtaagtcaaa gtcagatgat     300
ttccgtaaaa cgaatgctcg tggtcgaggt tatcagccta tggtttatgg actcaaatct     360
gtacggatta cagagcacaa tcaactggtt cgctttcatt tccagttcca aaaaggctta     420
gaaagggagt tcatctatcg tgtggaaaaa gaaaaaagt                           459
```

SeqID 86
```
atgaaaaaaa tgatgacatt cttgaaaaaa gctaaggtta aagcttttac attggtggag      60
atgttggtgg tcttgctgat tatcagcgtg cttttcttgc tctttgtacc taatctgacc     120
aagcaaaaag aagcagtcaa tgacaaagga aaagcagctg ttgttaaggt ggtggaaagc     180
caggcagaac tttatagctt agaaaagaat gaagatgcta gcctaagaaa gttacaagca     240
gatggacgca tcacggaaga acaggctaaa gcttataaag aatacaatga taaaaatgga     300
ggagcaaatc gtaaagtcaa tgat                                           324
```

SeqID 87
```
atgacatcaa aagttagaaa ggcagtcatc cctgctgctg gactaggaac tcgattttta      60
ccagcaacca aggcccttgc caaagaaatg ttgccaatcg tagacaaacc aactatccag     120
tttatcgtgg aagaagctct caaatcaggt attgaagata ttctagttgt cactggtaaa     180
tcaaaacgtt ctattgaaga ccacttgat tcaaacttcg aattggaata taacctcaaa     240
gaaaaaggga aaacagatct tttgaagcta gttgataaaa caactgacat gcgtctgcat     300
tttatccgcc aaactcatcc acgcggtctc ggagatgctg ttttgcaagc caaggctttc     360
gtcggaaatg aaccttttgt cgttatgctt ggtgatgact tgatggatat cacagacgaa     420
aaggctgttc cacttaccaa acaactcatg gatgactacg agcgtaccca cgcgtctact     480
atcgctgtca tgccagtccc tcatgacgaa gtatctgctt acggggttat tgctccgcaa     540
ggcgaaggaa aagatggtct ttacagtgtt gaaacctttg ttgaaaaacc agctccagag     600
gacgctccta gcgaccttgc tattatcgga cgctacctcc tcacgcctga aatttttgag     660
attctcgaaa agcaagctcc aggtgcagga atgaaattc agctgacaga tgcaatcgac     720
accctcaata aaacacaacg tgtatttgct cgtgagttca aggggctcg ttacgatgtc      780
ggagacaagt ttggcttcat gaaaacatcc atcgactacg ccctcaaaca cccacaagtc     840
aaagatgatt tgaagaatta cctcatccaa cttggaaaag aattgactga gaaggaa       897
```

SeqID 88
```
atgcaaatc aattaaatga attaaaacga aaaatgctgg aattttteca gcaaaaacaa       60
aaaaataaaa aatcagctag acctggcaag aaaggttcaa gtaccaaaaa atctaaaacc     120
ttagataagt cagtcatttt cccagctatt ttactgagta taaaagcctt atttaactta     180
```

```
ctctttgtac tcggttttct aggaggaatg ttgggagctg ggattgcttt gggatacgga    240
gtggccttat ttgacaaggt tcgggtgcct cagacagaag aattggtgaa tcaggtcaag    300
gacatctctt ctatttcaga gattacctat tcggacggga cggtgattgc ttccatagag    360
agtgatttgt tgcgcacttc tatctcatct gagcaaattt cggaaaatct gaagaaggct    420
atcattgcga cagaagatga acactttaaa gaacataagg gtgtagtacc caaggcggtg    480
attcgtgcga ccttggggaa atttgtaggt ttgggttcct ctagtggggg ttcaaccttg    540
acccagcaac taattaaaca gcaggtggtt ggggatgcgc cgaccttggc tcgtaaggcg    600
gcagagattg tggatgctct tgccttggaa cgcgccatga ataaagatga gattttaacg    660
acctatctca atgtggctcc ctttggccga aataataagg gacagaatat tgcaggggct    720
cggcaagcag ctgagggaat tttcggtgta gatgccagtc agttgactgt tcctcaagca    780
gcattttag caggacttcc acagagtccc attacttact ctccttatga aaatactggg    840
gagttgaaga gtgatgaaga cctagaaatt ggcttaagac gggctaaggc agttctttac    900
agtatgtatc gtacaggtgc attaagcaaa gacgagtatt ctcagtacaa ggattatgac    960
cttaaacagg acttttacc atcgggcacg gttacaggaa tttcacgaga ctatttatac   1020
tttacaactt tggcagaagc tcaagaacgt atgtatgact atctagctca gagagacaat   1080
gtctccgcta aggagttgaa aaatgaggca actcagaagt tttatcgaga tttggcagcc   1140
aaggaaattg aaatggtgg ttataagatt actactacca tagatcagaa aattcattct   1200
gccatgcaaa gtgcggttgc tgattatggc tatctttag acgatggaac aggtcgtgta   1260
gaagtaggga atgtcttgat ggataaccaa acaggtgcta ttctaggctt tgtaggtggt   1320
cgtaattatc aagaaaatca aaataatcat gcctttgata ccaaacgttc gccagcttct   1380
actaccaagc ccttgctggc ctacggtatt gctattgacc agggcttgat gggaagtgaa   1440
acgattctat ctaactatcc aacaaacttt gctaatggca atccgattat gtatgctaat   1500
agcaagggaa caggaatgat gaccttggga gaagctctga actattcatg gaatatccct   1560
gcttactgga cctatcgtat gctccgtgaa aagggtgttg atgtcaaggg ttatatggaa   1620
aagatgggtt acgagattcc tgagtacggt attgagagct gccaatggg tggtggtatt   1680
gaagtcacag ttgcccagca taccagacct gatcagacct tagctaataa tggagtttat   1740
catcagaagc atgtgatttc aaagattgaa gcagcagatg gtagagtggt gtatgagtat   1800
caggataaac cggttcaagt ctattcaaaa gctactgcga cgattatgca gggattgcta   1860
cgagaagttc tatcctctcg tgtgacaaca accttcaagt ctaacctgac ttctttaaat   1920
cctactctgg ctaatgcaga ttggattggg aagactggta caaccaacca agacgaaaat   1980
atgtggctca tgctttcgac acctagatta accctaggtg gctggattgg gcatgatgat   2040
aatcattcat tgtcacgtag agcaggttat tctaataact ctaattacat ggctcatctg   2100
gtaaatgcga ttcagcaagc ttccccaagc atttgggggga acgagcgctt tgctttagat   2160
cctagtgtag tgaaatcgga agtcttgaaa tcaacaggtc aaaaaccaga gaaggtttct   2220
gttgaaggaa aagaagtaga ggtcacaggt tcgactgtta ccagctattg ggctaataag   2280
tcaggagcgc cagcgacaag ttatcgcttt gctattggcg gaagtgatgc ggattatcag   2340
aatgcttggt ctagtattgt ggggagtcta ccaactccat ccagctccag cagttcaagt   2400
agtagttcta gcgatagcag taactcaagt actacacgac cttcttcttc aagggcgaga   2460
cga                                                              2463
```

SeqID 89

```
atgtcatcta aatttatgaa gagcgctgcg gtgcttggaa ctgctacact tgctagcttg     60
cttttggtag cttgcggaag caaaactgct gataagcctg ctgattctgg ttcatctgaa    120
gtcaaagaac tcactgtata tgtagacgag ggatataaga gctatattga agaggttgct    180
aaagcttatg aaaaagaagc tggagtaaaa gtcactctta aaactggtga tgctctagga    240
ggtcttgata aacttttctct tgacaaccaa tctggtaatg tccctgatgt tatgatggct    300
ccatacgacc gtgtaggtag ccttggttct gacggacaac tttcagaagt gaaattgagc    360
gatggtgcta aaacagacga cacaactaaa tctcttgtaa cagctgctaa tggtaaagtt    420
tacggtgctc ctgccgttat cgagtcactt gttatgtact acaacaaaga cttggtgaaa    480
gatgctccaa aaacatttgc tgacttggaa aaccttgcta agatagcaa atacgcattc    540
gctggtgaag atggtaaaac tactgccttc ctagctgact ggacaaactt ctactataca    600
tatggacttc ttgccggtaa cggtgcttac gtctttggcc aaaacggtaa agacgctaaa    660
gacatcggtc ttgcaaacga cggttctatc gtaggtatca actacgctaa atcttggtac    720
gaaaaatggc ctaaaggtat gcaagataca gaaggtgctg gaaacttaat ccaaactcaa    780
ttccaagaag gtaaaacagc tgctatcatc gacggacctt ggaaagctca gcctttaaa    840
gatgctaaag taaactacgg agttgcaact atcccaactc ttccaaatgg aaaagaatat    900
gctgcattcg gtggtggtaa agcttgggtc attcctcaag ccgttaagaa ccttgaagct    960
tctcaaaaat ttgtagactt ccttgttgca actgaacaac aaaaagtatt atatgataag   1020
actaacgaaa tcccagctaa tactgaggct cgttcatacg ctgaaggtaa aacgatgag   1080
ttgacaacag ctgttatcaa acagttcaag aacactcaac cactgccaaa catctctcaa   1140
atgtctgcag tttgggatcc agcgaaaaat atgctctttg atgctgtaag tggtcaaaaa   1200
gatgctaaaa cagctgctaa cgatgctgta acattgatca agaaacaat caaacaaaaa   1260
tttggtgaa                                                          1269
```

SeqID 90

```
atgatagata aagtggtcag gaacctactc ctgacctttt tcttttgcaa aatgacaaaa     60
atcataattt ttttgacaac tatacttgtc aaaaagaaaa agatatgtta caatgaattc    120
aagttaagaa ataggaagca gaaaggagtt ataatgtggg tactaggatt tatactattt    180
```

```
atgattttct tttattctaa taattctaaa aaaatcaaga aactagagaa taaaatcaaa        240
agacttgagc gaaaagagaa aggaaacgca gaaatgtcga gattattaca agaaatgatt        300
ggaaaggaac caattataac gggagtgtat attgggccag ataactggga agttgtggat        360
gttgatgagg aatgggtaaa gctacgacgt gtagataata cgggaaaaga aaaattcaag        420
ttgcaacgta ttgaggatat ccaaaccgtt gaatttgacg gagag                        465
```

SeqID 91
```
atgatttaa gtaaaaatag agaagatggg ttaagaaaat ttgcgactaa catccgatta         60
aatactctta gaacattgaa tcatcttgga ttcggacatt acggaggaag tctgtctatc        120
gtagaagttt tagcggtgct ttatggtgaa ataatgccaa tgactccaga aatatttgca        180
gcacgagata gagattattt catattatca aaaggtcacg gaggaccagc tctatacagt        240
acactctatt tgaatggttt ctttgacaaa gaattcttat attctttaaa tacaaatgga        300
accaaattac cgtctcatcc tgatagaaat ctaacgccag gcatagatat gacaacgggc        360
tctttaggac aaggaattag tgttgcaact ggacttgcat atggtcagag aataagaaag        420
agtccctttt atacttacgc tattgttgga gatggtgagt taaatgaggg acaatgttgg        480
gaggctatac agtttgcttc tcatcaacag ttatccaact taattgtatt tgttgatgat        540
aacaaaaaac aattagatgg ttttacaaag gatatttgta atccaggtga tttcgtagaa        600
aaattttcag catttggatt tgaatccatt agggtcaagg gttcagatat tagagaaatt        660
tatgaaggga ttgtccaatt aaaacagtca aataattcat cacctaagtg cattgtatta        720
gatactatta aaggtcaagg ggttcaagag ctggaagaaa tgaaatccaa tcatcatctt        780
cgccctactg tagaggagaa acaaatgtta acttcagttg tagaaagatt aagtcaggaa        840
ttggaggaaa cagaa                                                         855
```

SeqID 92
```
atgaaaaaaa ctacaatatt atcattaact acagctgcgg ttattttagc agcatatgtc        60
cctaatgaac caatcctagc agatactcct agttcggaag taatcaaaga gactaaagtt        120
ggaagtatta ttcaacaaaa taatatcaaa tataaggttc taactgtaga aggtaacata        180
ggaactgttc aagtgggtaa tggagttact cctgtagagt ttgaagctgg tcaagatgga       240
aaaccattca cgattcctac aaaaatcaca gtaggtgata aagtatttac cgttactgaa        300
gtagctagtc aagcttttag ttattatcca gatgaaacag gtagaattgt ctactatcct        360
agctctatta ctatcccatc aagcataaaa aaaatacaaa aaaaaggctt ccatggaagt        420
aaagctaaaa ctattatttt tgacaaaggc agtcagctgg agaaaattga agatagagct        480
tttgattttt ctgaattaga agagattgaa ttgcctgcat ctctagaata tattggaaca        540
agtgcatttt cttttagtca aaaattgaaa aagctaacct tttcctcaag ttcaaaatta        600
gaattaatat cacatgaggc ttttgctaat ttatcaaatt tagagaaact aacattacca        660
aaatcggtta aaacattagg aagtaatcta tttagactca ctactagctt aaaacatgtt        720
gatgttgaag aaggaaatga atcgtttgcc tcagttgatg gtgtttttgtt ttcaaaagat        780
aaaacccaat taatttatta tccaagtcaa aaaaatgacg aaagttataa aacgcctaag        840
gagacaaaag aacttgcatc atattcgttt aataaaaatt cttacttgaa aaaactcgaa        900
ttgaatgaag gtttagaaaa aatcggtact tttgcatttg cagatgcgat taaacttgaa        960
gaaattagct taccaaatag tttagaaact attgaacgtt tagccttta cggtaattta       1020
gaattaaaag aacttatatt accagataat gttaaaaatt ttggtaaaca cgttatgaac       1080
ggtttaccaa aattaaaaag tttaacaatt ggtaataata tcaactcatt gccgtccttc       1140
ttcctaagtg gcgtcttaga ttcattaaag gaaattcata ttaagaataa aagtacagag       1200
ttttctgtga aaaaagatac atttgcaatt cctgaaactg ttaagttcta tgtaacatca       1260
gaacatataa aagatgttct taaatcaaat ttatctacta gtaatgatat cattgttgaa       1320
aaagtagata atataaaaca agaaactgat gtagctaaac ctaaaaagaa ttctaatcag       1380
ggagtagttg gttgggttaa agacaaaggt ttatggtatt acttaaacga atcaggttca       1440
atggctactg gttgggttaa agacaaaggt ttatggtatt acttaaacga atcaggttca       1500
atggctactg gttgggttaa agacaaaggc ttatggtact acttaaatga atcaggttca       1560
atggctactg gttggggttaa agacaaaggc ttatggtatt acttaaacga atcaggttca       1620
atggctactg gttggggttaa agacaaaggc ttatggtact acttaaatga atcaggttca       1680
atggctactg gttgggttaa agacaaaggc ttatggtatt acttaaatga atcaggttca       1740
atggctactg gttgggttac agtttctggt aaatggtact ataacctataa ttcaggagat      1800
ttattagtaa acacgactac acccgatggc tatcgagtca atgctaacgg tgagtgggta       1860
gga                                                                     1863
```

SeqID 93
```
atggtaagat ttacaggact tagtctcaaa caaacgcaag ctattgaggt tttaaaaggt        60
cacatttctc taccagatgt ggaagtggct gtcactcagt ctgaccaagc atctatctct        120
atcgagggtg aggaaggtca ctatcaattg acctaccgca aacctcacca acttttatcgt        180
gccttgtcct tgttggtaac agttctagca gaagctgca agtagagat tgaggaacaa         240
gcagcttacg aagatttggc ttacatggtt gactgttctc gaaatgcggt gctgaatgtg        300
gcttctgcca agcagatgat tgagatattg gctctcatgg ctactcaac ctttgagctt        360
tacatggaag acacttacca gattgaaggg cagccttact ttggctattt ccgtggagct        420
tattcagcag aggagttgca ggaaatcgaa gcctatgccc aacagtttga cgtgacctttt        480
gtaccatgca tccagacctt ggcccacttg tcggcctttg tcaaatgggg tgtcaaggaa        540
gtgcaggagc tccgtgatgt agaggacatt cttctcattg gcgaagaaaa ggtttatgac        600
```

```
ttgattgatg gcatgtttgc cacgttgtct aaactgaaga ctcgcaaggt caatatcggg        660
atggacgaag cccacttggt tggtttggga cgctacctga ttctgaacgg tgttgtggat        720
cgtagtctcc tcatgtgcca acacttggag cgcgtgctgg atattgctga caaatatggt        780
ttccactgcc agatgtggag tgatatgttc ttcaaactca tgtcagcgga tggccagtac        840
gaccgtgatg tggaaattcc agaggaaact cgtgtctacc tagaccgtct caaagaccgt        900
gtgactctgg tttactggga ttattatcag gatagcgagg aaaaatacaa ccgtaatttc        960
cgcaatcatc acaagattag ccatgacctt gcatttgcag ggggagcttg gaagtggatt       1020
ggctttacac ctcacaacca ttttagccgt ctagtggcta tcgaggctaa taaagcctgc       1080
cgtgccaatc agattaaaga agtcatcgta acgggttggg gagacaatgg tggtgaaact       1140
gcccagttct ctatcctacc aagcttgcaa atctgggcag aactcagcta tcgcaatgac       1200
ctagatggtt tgtctgcgca cttcaagacc aatactggtc taacggttga ggattttatg       1260
cagattgacc ttgccaacct cttaccagac ctaccaggca atctcagcgg tatcaatccc       1320
aaccgctatg ttttttatca ggatattctt tgtccgattc ttgatcaaca catgacacct      . 1380
gaacaggaca aaccgcactt cgctcaggct gctgagacgc ttgctaacat taaagaaaaa       1440
gctggaaact atgcctatct ctttgaaact caggcccagt tgaatgctat tttaagtagc       1500
aaagtagatg tgggacgacg cattcgtcag gcctaccaag cggatgataa agaaagttta       1560
caacaaatcg ccagacaaga attaccagaa cttagaagcc aaattgaagc cttccatgcc       1620
ctctttagcc accaatggct gaaagaaaac aaggtctttg gtttggatac agttgacatc       1680
cgtatgggcg gactcttgca acgcatcaaa cgagcagaaa gccgtatcga ggtttatctg       1740
gctggtcagc ttgaccgcat cgacgagctg gaagttgaaa tcctaccatt tactgacttc       1800
tacgcagaca aggatttcgc agcaactaca gccaaccagt ggcataccat tgcgacagcg       1860
tcgacgattt atacgact                                                      1878
```

SeqID 94

```
atgtctaatt catttgtcaa gttgttagtc tctcaattat ttgcaaattt agcagatatt         60
ttctttagag taacaatcat tgctaacata tacattattt caaaatcagt aattgccaca        120
tcactagttc ctatcttaat aggaatatcc tcttttgttg cgagtctttt agttccgttg        180
gttactaaaa ggttagcgct aaataggggtt ttatctttat ctcaatttgg aaaagactata       240
ttattggcga tactggtagg aatgtttacc gtaatgcaat ccgtagcgcc tttggtgacc        300
tatctatttg ttgttgcaat ttccatacta gatggttttg cagcaccgt ttcctatgct          360
attgtgccac gctatgcgac cgatttgggt aaggctaatt cagccttatc aatgactggt        420
gaagctgttc aattgatagg ttggggatta ggtggactct tgtttgcaac aattggtctg        480
ttacctacca cgtgtatcaa tttagtcttg tatatcattt ctagctttct gatgttattt        540
cttcctaacg ctgaagtgga ggtgttagag tcagaaacta atcttgaaat tttgctcaaa        600
ggttggaagt tagttgctag aaatcctaga ttaagacttt ttgtatcagc aaatttattg        660
gaaattttt caaatacgat ttgggtttct tccattatac ttgttttgt aacggagtta        720
ttaaataaaa cggaaagtta ctggggatat tctaatacag catactctat tggtattata        780
attagtggct taattgcttt taggctatct gaaaagttcc ttgctgctaa a                  831
```

SeqID 95

```
atgtttgcat caaaaagcga aagaaaagta cattattcaa ttcgtaaatt tagtgttgga         60
gtagctagtg tagttgttgc cagtcttgtt atgggaagtg tggttcatgc gacagagaac        120
gagggagcta cccaagtacc cacttcttct aatagggcaa atgaaagtca ggcagaacaa        180
ggagaacaac ctaaaaaact cgattcagaa cgagataagg caaggaaaga ggtcgaggaa        240
tatgtaaaaa aaatagtggg tgagagctat gcaaatcacg ctaaaaagcg acatacaatt        300
actgtagctc tagttaacga gttgaacaac attaagaacg agtatttgaa taaaatagtt        360
gaatcaacct cagaaagcca actacagata ctgatgatgg agagtcgatc aaaagtagat        420
gaagctgtgt ctaagtttga aaaggactca tcttcttcgt caagttcaga ctcttccact        480
aaaccggaag cttcagatac agcgaagcca aacaagccga cagaaccagg agaaaaggta .       540
gcagaagcta agaagaaggt tgaagaagct gagaaaaaag ccaaggatca aaaagaagaa        600
gatcgtcgta actacccaac cattacttac aaaacgcttg aacttgaaat tgctgagtcc        660
gatgtggaag ttaaaaaagc ggagcttgaa ctagtaaaag tgaaagctaa cgaacctcga        720
gacgagcaaa aaattaagca agcaaagcg gaagttgaga gtaaacaagc tgaggctaca .       780
aggttaaaaa aaatcaagac agatcgtgaa gagcagaag aagaagctaa acgaagagca        840
gatgctaaag agcaaggtaa accaaagggg cgggcaaaac gaggagttcc tggagagcta        900
gcaacacctg ataaaaaga aaatgatgcg aagtcttcag attctagcgt aggtgaagaa        960
actcttccaa gcccatccct gaaaccagaa aaaaaggtag cagaagctga gaagaaggtt       1020
gaagaagcta agaaaaaagc cgaggatcaa aaagaagaag atcgccgtaa ctacccaacc       1080
aatacttaca aaacgcttga acttgaaatt gctgagtccg atgtggaagt taaaaaagcg       1140
gagcttgaac tagtaaaaga ggaagctaag gaacctcgaa acgaggaaaa agttaagcaa       1200
gcaaaagcgg aagttgagag taaaaaagct gaggctacaa ggttagaaaa aatcaagaca       1260
gatcgtaaaa aagcgaac agagaagcta aagaagctaa acgaagagca taaagttaaa        1320
gaaaaccag ctgaacaacc acaaccagcc ccggctccaa aagcagaaaa accagctcca       1380
gctccaaaac cagagaatcc agctgaacaa ccaaaagcag aaaaaccagc tgatcaacaa       1440
gctgaagaag actatgctcg tagatcagaa gagaatatat cgcttgac tcaacagcaa        1500
ccgccaaaaa ctgaaaaacc agcacaacca tctactccaa aaacaggctg aaacaagaa        1560
aacggtatgt ggtacttcta caatactgat ggttcaatgg cgacaggatg gctccaaaac       1620
aatggctcat ggtactacct caacagcaat ggcgctatgg cgacaggatg gctccaaaac       1680
```

```
aatggttcat ggtactatct aaacgctaat ggttcaatgg caacaggatg gctccaaaac   1740
aatggttcat ggtactacct aaacgctaat ggttcaatgg cgacaggatg gctccaatac   1800
aatggctcat ggtactacct aaacgctaat ggttcaatgg cgacaggatg gctccaatac   1860
aatggctcat ggtactacct aaacgctaat ggtgatatgg cgacaggttg ggtgaaagat   1920
ggagatacct ggtactatct tgaagcatca ggtgctatga aagcaagcca atggttcaaa   1980
gtatcagata aatggtacta tgtcaatggc tcaggtgccc ttgcagtcaa cacaactgta   2040
gatggctatg gagtcaatgc caatggtgaa tgggtaaac                          2079
```

SeqID 96
```
atgaactatt caaaagcatt gaatgaatgt atcgaaagtg cctacatggt tgctggacat     60
tttggagctc gttatctaga gtcgtggcac ttgttgattg ccatgtctaa tcacagttat    120
agtgtagcag gggcaacttt aaatgattat ccgtatgaga tggaccgttt agaagaggtg    180
gctttggaac tgactgaaac ggactatagc caggatgaaa cctttacgga attgccgttc    240
tcccgtcgtt tgcaggttct ttttgatgaa gcagagtatg tagcgtcagt ggtccatgct    300
aaggtactag ggacagagca cgtcctctat gcgattttgc atgatagcaa tgccttggcg    360
actcgtatct tggagagggc tggtttttct tatgaagaca agaaagatca ggtcaagatt    420
gctgctcttc gtcgaaattt agaagaacgg gcaggctgga ctcgtgaaga tctcaaggct    480
ttacgccaac gccatcgtac agtagctgac aagcaaaatt ctatggccaa tatgatgggc    540
atgccgcaga ctcctagtgg tggtctcgag gattatacgc atgatttgac agagcaagcg    600
cgttctggca agttagaacc agtcatcggt cgggacaagg aaatctcacg tatgattcaa    660
atcttgagcc ggaagactaa gaacaaccct gtcttggttg gggatgctgg tgtcgggaaa    720
acagctctgg cgcttggtct tgcccagcgt attgctagtg gtgacgtgcc tgcggaaatg    780
gctaagatgc gcgtgttaga acttgatttg atgaatgtcg ttgcaggcac acgcttccgt    840
ggtgactttg aagaacgcat gaataatatc atcaaggata ttgaagaaga tggccaagtc    900
atcctcttta tcgatgaact ccacaccatc atgggttctg tagcgggat tgattcgact    960
ctggatgcgg ccaatatctt gaaaccagcc ttggcgcgtg gaactttgag aacggttggt   1020
gccactactc aggaagaata tcaaaaacat atcgaaaaag atgcggcact ttctcgtcgt   1080
ttcgctaaag tgacgattga agaaccaagt gtggcagata gtatgactat tttacaaggt   1140
ttgaaggcga cttatgagaa acatcaccgt gtacaaatca cagatgaagc ggttgaaaca   1200
gcggttaaga tggctcatcg ttatttaacc agtcgtcact tgccagactc tgctatcgat   1260
ctcttggatg aggcggcagc aacagtgcaa aataaggcaa agcatgtaaa agcagacgat   1320
tcagatttga gtccagctga caaggccctg atggatggca agtggaaaca ggcagcccag   1380
ctaatcgcaa aagaagagca agtacctgtc tggtgacaga gtctgatatt               1440
ttgaccacct tgagtcgctt gtcaggaatc ccagttcaaa aactgactca aacggatgct   1500
aagaagtatt taaatcttga agcagaactc cataaacggg ttatcggtca agatcaagct   1560
gtttcaagca ttagccgtgc cattcgccgc aaccagtcag ggattcgcag tcataagcgt   1620
ccgattggtt cctttatgtt cctagggcct acaggtgtcg ggaaaactga attagccaag   1680
gctctggcag aagttctttt tgacgacgaa tcagccctta tccgctttga tatgagtgag   1740
tatatggaga aatttgcagc tagtcgtctc aacggagctc ctccaggcta tgtaggatat   1800
gaagaaggtg gggagttgac agagaaggtt cgcaataaac cctattccgt tctcctcttt   1860
gatgaggtag agaaggcccca cccagatatc tttaatgttc tcttgcaggt tctggatgac   1920
ggtgtcttga cagatagcaa gggacgcaag gtcgattttt caaataccat tatcattatg   1980
acatcgaatc taggtgcgac tgcccttcgt gatgataaga ctgttggttt tggggctaag   2040
gatattcgtt ttgaccagga aaatatggaa aaacgcatgt ttgaagaact gaaaaaagct   2100
tatagaccgg aattcatcaa ccgtattgat gagaaggtgg tcttccatag cctatctagt   2160
gatcatatgc aggaagtggt gaagattatg gtcaagcctt agtggcaag tttgactgaa   2220
aaaggcattg acttgaaatt acaagcttca gctctgaaat tgttagcaaa tcaaggatat   2280
gacccagaga tgggagctcg cccacttcgc agaaccctgc aaacagaagt ggaggacaag   2340
ttggcagaac ttcttctcaa gggagattta gtggcaggca gcacacttaa gattggtgtc   2400
aaagcaggcc agttaaaatt tgatattgca                                    2430
```

SeqID 97
```
atgaaaattt taccgtttat agcaagagga acaagttatt acttgaagat gtcagttaaa     60
aagcttgttc ctttttagt agtaggattg atgctagcag ctggtgatag tgtctatgcc    120
tattccagag gaaatggatc gattgcgcgt ggggatgatt atcctgctta ttataaaaat    180
gggagccagg agattgatca gtggcgcatg tattctcgtc agtgtacttc ttttgtagcc    240
tttcgtttga gtaatgtcaa tggttttgaa attccggcag cttatggaaa tgcgaatgaa    300
tggggacatc gtgctcgtcg ggaaggttat cgtgtagata atacaccgac gattggttcc    360
attacttggt ctactgcagg aacttatgct catgttgcct gggtgtcaaa tgtaatggga    420
gatcagattg agattgagga atataactat ggttatacag aatcctataa taaacgagtt    480
ataaaagcaa acacgatgac aggatttatt cattttaaag atttggatgg tggcagtgtt    540
gggaatagtc aatcctcaac ttcaacaggc ggaactcatt attttaagac caagtctgct    600
attaaaactg aacctctagc tagcggaact gtgattgatt actattatcc tggggagaag    660
gttcattatg atcagatact tgaaaaagac ggctataagt ggttgagtta tactgcctat    720
aatggaagct atcgttatgt tcaattggag ctgtgaata aaaatcctct aggtaattct    780
gttctttctt caacaggtgg aactcattat tttaagacca agtctgctat caaaactgaa    840
cccctagtta gtgcaactgt gattgattac tattatcctg agagaaggt tcattatgat    900
caaattctcg aaaaagacgg ctacaagtgg ttgagttata cggcttataa cggaagtcgt    960
```

```
cgctatatac agctagaggg agtgacttct tcacaaaatt atcagaatca atcaggaaac   1020
atctctagct atggatccca tagtagttca actgtcggtt ggaagaaaat aaatggtagt   1080
tggtatcatt tcaaatcaaa tggttctaaa tcaacaggat ggctgaaaga cggttctagc   1140
tggtattatt tgaaattatc tggtgaaatg cagacaggat ggttaaagga aaatggtttg   1200
tggtattatc tgggtagttc aggggcaatg aaaacaggct ggtaccaggt ctctggtaag   1260
tggtattatt cttactcttc aggcgcctta gctgttaata cgacggtgga tggctacaga   1320
gtaaacagtg atggagaacg agta                                          1344
```

SeqID 98
```
atgaaagtaa tcttttagc agatgttaaa ggaaaaggta aaaaaggcga aattaaggaa    60
gtaccaacag ggtatgcgca aaactttctt atcaaaaaga atctagccaa agaagcgact   120
gctcaagctg taggtgaact tcgtggtaaa caaaaatcgg aagaaaaagc tcacgctgag   180
atgattgcag aaggaaaagc aattaaagca caacttgaag cagaagaaac tgttgtagaa   240
tttgttgaaa aagttggtcc agatggtcgt acctttggtt ctattaccaa taagaagatt   300
gcagaagaat tgcaaaagca atttggaatt aagattgata aacgtcatat tcaagtacaa   360
gctccgattc gagcggttgg ttttgattgat gtgccagtga aaatctatca agatatcaca   420
agtgtaatca atcttcgtgt gaaagaagga                                    450
```

SeqID 99
```
atgaagaaaa aaatcttagc gtcactttta ttaagtacag taatggtttc tcaagtagct    60
gttttaacaa ctgcgcatgc agaaacgact gatgacaaaa ttgctgctca agataataaa   120
attagtaact taacagcaca acaacaagaa gcccaaaaac aagttgacca aattcaggag   180
caagtatcag ctattcaagc tgagcagtct aacttgcaag ctgaaaatga tagattacaa   240
gcagaatcta agaaactcga gggtgagatt acagaacttt ctaaaaacat tgtttctcgt   300
aaccaatcgt tggaaaaaca agctcgtagt gctcaaacaa atggagccgt aactagctat   360
atcaatacca ttgtaaactc aaaatcaatt acagaagcta tttcacgtgt tgctgcaatg   420
agtgaaatcg tatctgcaaa caacaaaatg ttagaacaac aaaaggcaga taaaaaagct   480
atttctgaaa aacaagtagc aaataatgat gctatcaata ctgtaattgc taatcaacaa   540
aaattggctg atgatgctca agcattgact acgaaacagg cagaactaaa agctgctgaa   600
ttaagtcttg ctgctgagaa agcgacagct gaagggga aagcaagtct attagagcaa     660
aaagcagcag ctgaggcaga ggctcgtgca gctgcggtag cagaagcagc ttataaagaa   720
aaacgagcta gccaacaaca atcagtactt gcttcagcaa acactaactt aacagctcaa   780
gtgcaagcag tatctgaatc tgcagcagca cctgtccgtg caaaagttcg tccaacatac   840
agtacaaacg cttcaagtta tccaattgga gaatgtacat ggggagtaaa aacattggca   900
ccttgggctg gagactactg gggtaatgga gcacagtggg ctacaagtgc agcagcagca   960
ggtttccgta caggttcaac acctcaagtt ggagcaattg catgttggaa tgatggtgga   1020
tatggtcacg tagcggttgt tacagctgtt gaatcaacaa cacgtatcca agtatcagaa   1080
tcaaattatg caggtaatcg tacaattgga aatcaccgtg atggttcaa tccaacaaca     1140
acttctgaag gttttgttac atatatttat gcagat                            1176
```

SeqID 100
```
atggtaaaaa gacgtataag gagagggacg agagaacctg aaaaagttgt tgttcctgag    60
caatcatcta ttccttcgta tcctgtatct gttacatcta accaaggaac agatgtagca   120
gtagaaccag ctaaagcagt tgctccaaca acagactgga aacaagaaaa tggtatgtgg   180
tattttttata atactgatgg ttccatggca acaggttggg tacaagttaa tagttcatgg   240
tactacctca acagcaacgg ttctatgaaa gtcaatcaat ggttccaagt tggtggtaaa   300
tggtattatg taaatacatc gggtgagtta gcggtcaata caagtataga tggctataga   360
gtcaatgata atggtgaatg ggtgcgt                                       387
```

SeqID 101
```
gagttgcgac ggctatcaag gttggtggac caggagctct attttggatg tggatggcgg    60
ctttctttgg aatggctacc aagtatgcgg aaggactctt ggccatcaaa taccgcacca   120
aggacgacca tggtgcag                                                 138
```

SeqID 102
```
gactgtatca ggaaacaacc gttcacacgc gatgaaccaa acaaaacgtg ccgtaaaacc    60
aaaccttcaa aaagttactg ttcttatcga tgg                                93
```

SeqID 103
```
ggacaaagaa accctcgaag aattgaaaga gttatcagaa tggcagaaac gaaaccaaga    60
atatctaaaa aagaaggc                                                 78
```

SeqID 104
```
cagaggaagc tgttcaaaat cttccaccta ttccagaaga aaagtgggtg gaaccagaaa    60
tcatcctgcc tcaagctgaa cttaaattcc ctgaacagga agatgactca gatgacgaag   120
atgttcaggt cgattttca gccaaagaag cccttgaata caaacttcca agcttacaac     180
tctttgcacc agataaacca aaagatcagt ctaaagagaa gaaaattgtc agagaaaata   240
tcaaaatct                                                          249
```

```
SeqID 105
ttggtgatta tagttttgaa aatccagtcc aaatcggaga cagactttat tttcaagaca    60
tggccatttta ttcttttgtc aaaaataata cctttaatgg tattggattg ccaagtctct   120
atctcatgga cgaacaggga gactgtagct tactcaaagc ttttggctat caagacttta   180
aagggagatt atcatgatgg acagtccaaa aaaattaggc tatcacatgc cagcagagta   240
cgaacccat catggtaccc tcatgatatg gccgactcga ccaggatcat ggccttttca    300
aggaaaggct gc                                                        312


SeqID 106
gagagactac cagcttttcc tagaagtctt tcagggagga agttggacca aggcggaacc    60
aaagaaaaag gctcggatgg aagaagtcct                                      90


SeqID 107
agaaattgcc tctctacttg gaaaagctcc tcaaactatc acactgaaat caagcgtggg    60
acagtccgac aatgtcttgg aaaagggcgc ttcaaagagg tttattctgc cgactacgct   120
caacagtctt atgaaaacaa tcgcaagcgc tcggtcaaga aatcaagctt gaccaaggaa   180
ctaaaggaaa agattctcca ctatcataac caaaaatttt cgcctgaaat gatggttatg   240
gctaaagggg ttaacgtggg aatttcaacc atttactatt ggattcatca tggaaaattg   300
gggttaagca agcaggattt gctttaccct agaaaaggaa aagcgcttaa gaaacaggct   360
agcaccaact ttaaacctgc tggtcaatcc atcgaacagc ggcctgaagc tatcaatctt   420
cgcttggaga atgggcatta tgagattgat acggttctac ttacgagatc gaaaaactac   480
tgcttgattg tcttgacgga tcgaaagagt agacatcaga tcatccgatt gattccaaat   540
aaaagtgctg aggtggtcaa tcaggctcta aaactcatct taaaacaaca caagattctt   600
tccatcacgg cagataatgg aacggaattc aatcgcttgt ttgatatatt ttctgaggag   660
cacatctatt atgcgcaccc ctatgcctct tgggaaaggg gaactaatga gaatcacaac   720
aggctcattc gtagatagtt acctaaggga accaagaaaa tgactcccaa agaagtcgca   780
ttcatcgaaa agtggattaa caactatcct aaaaaatgct tggactacaa gtcacccaga   840
gaagacttct ggatggctaa cttgaacttg aaatttagca aaatggaaat aattttttatt  900
aaacgcttcc aa                                                        912


SeqID 108
cctgtcatga ctatctcatc gcctactatg aaaaacatgg atttgtcaac gaaggccagt    60
cccagtcaac ctttgcaggg gaaacatggt atgatatggt ctgggaaa                108


SeqID 109
acatcatcaa taaggataca cacaaggaaa tcatcgccaa actggactac gacgccccat    60
cttgccctga gtgcggaaac caat                                            84


SeqID 110
tacttccttc cacacaagta tgccagagaa agcttatcgc taccctctac caacaaaata    60
ttacacagaa aacaaggttc g                                               81


SeqID 111
gcagccttca aaaagatca aattaatgag cgtgtcgaga aattaggtaa gttaaaacct      60
attacaataa attacaacgg aaaatcagaa gtaattgata gtaaagaaaa attacaagag    120
cttatgaata aagccgttaa agacgaagtg gctcaaata                           159


SeqID 112
gcttatgcgc attctaaaag aagcgctgga agtggcaggg caggaggcag acaatgtctt     60
tgccaatgtc aaaataaatg taggagagat tttaagtat                            99


SeqID 113
cacggccgac cttatcataa gcctcaccaa ccgcatcatc acgggtttcc ccaacaatct     60
tataatctcc tgcctccgaa acataaacca actctgtgtg tccgccgc                 108


SeqID 114
aaaggaaaaa tcctcctgct accaaggcta accactcaa gatggcaaag aaaaatccgc      60
cctgactcac gtaagtcagc aaataataaa gcaaaccttg acttccataa tagtcgctgt    120
aaatcttccc tgtctgatga agcgcccaac ctgcataaaa atcctgcact tcttgtgcac    180
tcattaagtc gagtaatagc ggtactccta gagttatccc cgttacaagc gtactccata   240
gtaaaatttt caccaaagga agacgacttg attcacgatg atgcgattct tgttcgattt    300
ggtattctag aggttcacga ttctccttat gaacttcttc tactctacca tacacactca   360
tatcgtttct cctgttcaat ttatctgtct                                     390


SeqID 115
tttacggtaa gccatgtatt cctcctttat ttatctttta atccaagacc caaatcaatg     60
agtttgagtt tcacttcttc caaactcttg cgtccaagat ttcgtacttt catcatctct    120
```

```
gcttcagatt tttctgtcaa atcatgcaca gtattgatac cggcacgttt taaacagttg      180
tatgaacgca cagacaagtc cagttcctca atcgtacgat ctaaaatacg gtcgtcagat      240
tcagtatcag cttctttcat cacttcagtt gacttagcaa tctcagtaag atttgtaaac      300
aaaatcaagat gttctgtcaa aatacgtgct gaaagcccta aagcatcttc tggaataatt      360
gttccatttg tcaagatttc aagggttaat ttgtcgaaac catcattgct acctacacga      420
gcaggttcca ct                                                          432
```

```
SeqID 116
cttgtctgca tgaagaataa gggctgctac aaggaaagaa acaactgctg ccac             54
```

```
SeqID 117
ttccattatt tgtcaaaata cttttttagtt tcagcaataa cgactggcga caagaccaag      60
agggcaatca agtttggcag agccatcaag gcgttaacga tatctgcgat aatccagacc      120
atatccaact cgataaatcc tcctaacaag accatgagca caaaaaccac acgg            174
```

```
SeqID 118
ggaaagaagg tattcataaa atacctctca tcaagagtct cctcaaaaac aggaccgatg       60
attacaggca ggacaaaaga taagatagtc gataaaaagg ttggttgtcc atttgaaaaa      120
agcacggtaa aatactcatc a                                                141
```

```
SeqID 119
tcttcaccag ttttttcctaa acttgtaatg gtatctgggg caaataaacc aagagaaagg      60
cgcaatttcc cattttcgtc taaaatgtca ttccacttaa cctttgtctt g               111
```

```
SeqID 120
tacttaactt ccttctcagt tccgaagata gcttcttcaa aggtcaaatt gacacgatac       60
tggagatcat ctccttggcg aggagcgttt ggattgcgcg aagaaccgcc tccgccgaag      120
aaacttgaga aaatatcctc aaaaccaccg aagccacctg ccccattgaa accgccgaaa      180
ccaccagctc caccaaaacc accattggcg cctgcagcac catactggtc a               231
```

```
SeqID 121
cagtcatggc gtcctattcc agattcaaaa tgctatacac aagaaaaact cactatcccc       60
attaaaagaa gaaaagacat caaggacttc taccacaatt ccatccaaag acacaaaaac      120
agccataaga gtcacctcct tgattcctat aggctgatta taacaagact ggctgaaatt      180
gtacatgaaa ataaaatcct aatagtactc attttgtatg tgactaatat tccgtctcgc      240
tccagaaggt acgaagtaaa tagagtt                                          267
```

```
SeqID 122
ctgtttcgtt tttatcgtgt aattgttctt tatcgaggtt ggcatattta tcttttaatt      60
cttgtgaatt tgcagtacgt tcaaaacgtt ttccgaaagg atcgattcct tgttcgcgga      120
gcgcagccat tttttcacgg cgaacgatct gctggtcatt tagttcttcc atatgttctg      180
```

```
SeqID 123
ataactcacc ctccactaaa ccctgagcat tttgtttcaa gagtcttttc atctcttggt       60
ttgaagtctt atcagccaaa agatgataga tttctgagaa agccttcaga tagtaggcat      120
cctgaatcag gtaatagcgg aaaatggcag gttctaaatt cccctcttgt aattgtaaaa      180
```

```
SeqID 124
acattggcta aagcagtcgg tttgatgtat tctccaccaa ttccaccaaa accattctta       60
ggccgaataa cgacagattc gtcttctata                                       90
```

```
SeqID 125
cctggttctc cattttcaga gattccggt gcaggatttt ttggtgtcgc gaaacgaata        60
tttccacgtc caccacgacc accgtgggca acgataaatt cttgcccatg ttcaatcaaa      120
tctgttaaaa ccttgccagt ctccgcatca cgaacagtcg taccttgtgg tactcgaact      180
ctaaggtcct cagcaccacg accatgcatc cctttggtca tccctttttc accagaatca      240
gccttgaaa                                                              249
```

```
SeqID 126
ataaattcct tgaccttggc cacatcctta tccaaaagaa gggcaccaag aaaggcttca       60
aaggcatcac caagaatggt gtcacgattg cgaccacctg attttttcttc cccttttaccc     120
aacttgataa actggtcaaa ctggcaatca cgcgcaaaac cagctaaact ctcctcacgg      180
acaatcatag cacggagttt tgataggtca ccttcaggct tttttaggata tttttttatat    240
agatattctg aaatcaataa ctgtagaaca gcgtctccta aaaattccaa gcgttcattg      300
tgtgaaattt ttaagaggcg gtgctcattg gca                                   333
```

```
SeqID 127
ccaataggaa aaaggaattg taaagctgaa tgccaatccc accacctgct tgaaaagcag       60
```

```
aagaccttcc agtcaagaaa gaccaagaga tatggggcaa gccccgaacc aagatataga     120
gaatcaagga agccaagatt gtcacaa                                         147

SeqID 128
cagccattgg gacactcgaa agccgaagaa catgagacta tctgttcgca taccttcgat      60
aaccatacga ccgaaaccat accaaatcaa gtaaaaggcc gtgatatgac ctcgtctgag     120
actcttccat ttccgtctaa aaatcagaat caaggcaaag ccaagcagat tcca          174

SeqID 129
ccttgctctt tacctgatta tgggctggtt ggttctggct atcattcctg ccattatcag      60
tcaaacgaca cccgttttct ggagtctcat ggtaactggc ggactctgtt atacagttgg     120
agctggattt tatgccaaga aaaaacctta tttccacatg atttggcatc tctttatcct     180
agctgcgtcc gcacttcaat acatcgctat tgtttattac atgtaaaaaa gttgagaaat     240
tcaatctcaa cttttttctt tacacatatt gataaagtac tggtgcaagc gcacatcatc     300
agtcaattct ggatgaaaag aacttaccaa catatttttt tcttgggctg caacaatttg     360
attgttcact gttgc                                                     375

SeqID 130
cgagtaaaag ataatcatct ggataagctt gtgaaagctc ttctaaaaag gcgttcatcc      60
actcagtatt acatccacca gctattaaga aaaatgattc gcctgtatgg gcatcaacag     120
ctccataaca atagcgaaat tctcgtatat agtgactatg gacatgtgga cctactccta     180
ttggagacca acaagatccc agtttac                                        207

SeqID 131
caagtcatca aaatagacat agcaactaca aataaaacgg aatctgtaaa gagccaaagt      60
gagagagaaa agaaaagatt gacaagcagt aatatactaa aggttagagg gcgaccgata     120

SeqID 132
gcctttaaga gttccaaggt cccatcactt gatccatcat cgacaaagac atactcgatt      60
tctgtttcca aatctggaag taaagcttcc agagcc                               96

SeqID 133
gataaaactg acccactggc taggaaactt cctgacaaaa gtaagccgtc aacttccttt      60
tgcaccaaat cactttctcc cgttaacatg gcttcattga cttccgcaaa gccttccaaa     120
accaaggcat cactaggaat ctgctctcct gcagacaaac gaatgacatc tcctagcact     180
aattcttcag gattaagagc aacttcc                                        207

SeqID 134
ttatgccgat tacaaacaca agcaaggcca cgagggtctg tgaccaatct aacgaagcaa      60
aataaggtat atagatacct aaattatctc cgccagacgc aattgtcagc aatg          114

SeqID 135
tgcattcaaa gcattggcaa tgagggacag tgcaaaggca atagttgtta cgtaggcaag      60
gagattcatc ttgcccccat atccgatata gttggtcaca aaggcaaaga ggaaggcgat     120
gatggaaatg atgatggccg ccaattttac ctgtttttgg ctcattggt tgggtctgcc      180
ttcttgcgaa gcttcccact tctttatagc aaaggtataa atgaggaagg tgacgggata     240
ggtaatgatg gccgccttat ttccaaggat ataatcaata gcaccggaca aaatggtatt     300
aacaatacca agtaatttc cccatttgct                                      330

SeqID 136
acgtccacga agctggttat cgatacgacg actttcatga cgttctgtac caataacaca      60
aagtcctcca agttcacgaa caccttcacc aagcttgatg tcggtaccac gacccgccat     120
gttggttgcg atggtaacgg caccacgttg accagcattc atgatgattt gggcttctct     180
atagtggttt ttggcattca agacttcgtg aggaacacca gctgcaacca atttcttaga     240
aatgtagtca ctagtttcaa ccgctactgt accaaccaag acaggttgac ccttttggta     300
acgagcctta acgtcttcga caaccgcttt aaacttagat tcgatacttg cataaagaag     360
gtc                                                                  363

SeqID 137
atctctactg gtgtaccgac ctgttcgatg tatccattgt taaagactgc aattctatca      60
gataaagtca aggcttcctc t                                               81

SeqID 138
ttaagtacca tgtccagcat aaagtcaatc ttgtgctctt taccgacaca caccattttc      60
tcaaaatcag ccatatcacc aaaaagagga tccactgcca ta                       102

SeqID 139
agctgctcat actcatctac caactccaag gcatgctcaa tcgtcggttt atcaaaacca      60
```

```
acaatattca tctgtgtcac acccatctca gcagccaagg caatttcttc tgcctcacga     120
cctgcaaatt cttctacatt caagtgtgta tgtgtatcaa aaatcatctc ttctaacctc     180
gttttctatc ttctattata ccaaaaaaga ggaggggcac ctaatttttc ggtttcccct     240
cctctcttca atagagagct attctgctat cttttctatc cgatattgcc catctcctat     300
tccacagtta gagacagaag agattggcta cat                                 333
```

SeqID 140
```
gtaacatctt gcattgttcc tgcagttgct tgcggtgcac ttgttgtctt aggagcagca     60
ttaggagcca ctggtctctt aggaactgta accatggcaa tggcttgcac tccaatcgtt     120
tctgcttcct tcacatcttc catagttttc gcattttcaa tctctttgag agctgcttgt     180
ttttctgctt ccactcttgc taaaagttct gcttttctt tatcagaaag cggtgcgcct     240
ttgatttcat cctgcttgtc cttggctgct ttttcaatgg cattttttagc tgattctttt     300
tcagtagcta attgcttaga agcttgtgat                                     330
```

SeqID 141
```
tattctcctt tcaaccactc cattctcata aggaaaacga cgaaaatcat aaatccaaac     60
cccaaagcac cacgaatgaa ttggcgaagc aaggtttggt caaaccaacc tgtaaacatt     120
tccactaacc ataccaagag tgacaggccg ataaagaaa                           159
```

SeqID 142
```
gattatttca agtttcgaac aacttttaca agattttcta cagtaaagcc atattctgcc     60
aatactttg gtgctggggc agaggctccg aaggtatcaa tacctagaac ggcaccatcg     120
agaccaacat atttgtacca gttt                                           144
```

SeqID 143
```
agaggcagac gtggattatg cgttgcacga atcaaggctc ctagactagt cattaaacct     60
aagagaacaa tcgatccgcc taccaaagat agatacagtc caccactctc agctacatcc     120
ctctccgtcc ccaaaagtcc tatcatctct ttcccagcga agatggacaa aaatcctaaa     180
aggaaactta atagtaaggt aatcttcaac gcctcagtca ca                       222
```

SeqID 144
```
actcctccat ataccaaaat tcctgccaaa acagctataa taccatttat ttcagctcaa     60
gatttcaacc aagcccaacg gctctctgga                                     90
```


SeqID 145
MSKNIVQLNNSFIQNEYQRRRYLMKERQKRNRFMGGVLILIMLLFILPTFNLAQSYQQLLQRRQQLADLQTQYQTLSDEKDKE
TAFATKLKDEDYAAKYTRAKYYYSKSREKVYTIPDLLQR

SeqID 146
MDKKKLLLIDGSSVAFRAFFALYQQLDRFKNVAGLHTNAIYGFQLMLSHLLERVEPSHILVAFDAGKTTFRTEMYADYKGGRA
KTPDEFREQFPFIRELLDHMGIRHYDLAQYEADDIIGTLDKLAEQDGFDITIVSGDKDLIQLTDKHTVVEISKKGVAEFEAFI
PDYLMEEMGLTPAQFIDLKALMGDKSDNIPGVTKVGEKTGIKLLLEHGSLEGIYENIDGMKTSKMKENLINDKEQAFLSKTLA
TIDTKAPIAIGLEDLVYSGPDVENLGKFYDEMGFKQLKQALNVSSADVSESLDFTIVDQISQDMLSEESIFHFELFGENYHTD
NLVGFVWSCGDKLYATDKLELLQDPIFKDFLEKTSLRVYDFKKVKVLLQRFGVDLQAPAFDIRLAKYLLSTVEDNEIATIASI
YGQTYLVDDETFYGKGVKKAIPEREKFLEHLACKLAVLVETEPILLEKLSENGQLELLYDMEQPLAFVLAKMEIAGIMVKKET
LLEMQAENELVIEKLTQEIYELAGEEFNVNSPKQLGVLLFEKLGLPLEYTKKTKTGYSTAVDVLERLAPIAPIVKKILDYRQI
AKIQSTYVIGLQDWILADGKIHTRYVQDLTQTGRLSSVDPNLQNIPARLEQGRLIRKAFVPEWEDSVLLSSDYSQIELRVLAH
ISKDEHLIKAFQEGADIHTSTAMRVFGIERPDDVTANDRRNAKAVNFGVVYGISDFGLSNNLGISRKEAKAYIDTYFERFPGI
KNYMDEVVREARDKGYVETLFKRRRELPDINSRNFNIRGFAERTAINSPIQGSAADILKIAMIQLDKALVAGGYQTKMLLQVH
DEIVLEVPKSELVEMKKLVKQTMEEAIQLSVPLIADENEGATWYEAK

SeqID 147
MGMAAFKNPNNQYKAITIAQTLGDDASSEELAGRYGSAVQCTEVTASNLSTVKTKATVVEKPLKDFRASTSDQSGWVESNGKW
YFYESGDVKTGWVKTDGKWYYLNDLGVMQTGFVKFSGSWYYLSNSGAMFTGWGTDGSRWFYFDGSGAMKTGWYKENGTWYYLD
EAGIMKTGWFKVGPHWYYAYGSGALAVSTTTPDGYRVNGNGEWVN

SeqID 148
MSRKSIGEKRHSFSMRKLSVGLVSVTVSSFFLMSQGIQSVSADNMESPIHYKYMTEGKLTDEEKSLLVEALPQLAEESDDTYY
LVYRSQQFLPNTGFNPTVGTFLFTAGLSLLVLLVSKRENGKKRLVHFLLLTSMGVQLLPASAFGLTSQILSAYNSQLSIGVGE
HLPEPLKIEGYQYIGYIKTKKQDNTELSRTVDGKYSAQRDSQPNSTKTSDVVHSADLEWNQGQGKVSLQGEASGDDGLSEKSS
IAADNLSSNDSFASQVEQNPDHKGESVVRPTVPEQGNPVSATTVQSAEEEVLATTNDRPEYKLPLETKGTQEPGHEGEAAVRE
DLPVYTKPLETKGTQGPGHEGEAAVREEEPAYTEPLATKGTQEPGHEGKATVREETLEYTEPVATKGTQEPEHEGEAAVEEEL
PALEVTTRNRTEIQNIPYTTEEIQDPTLLKNRRKIERQGQAGTRTIQYEDYIVNGNVVETKEVSRTEVAPVNEVVKVGTLVKV
KPTVEITNLTKVENKKSITVSYNLIDTTSAYVSAKTQVPHGDKLVKEVDIENPAKEQVISGLDYYTPYTVKTHLTYNLGENNE
ENTETSTQDFQLEYKKIEIKDIDSVELYGKENDRYRRYLSLSEAPTDTAKYFVKVKSDRFKEMYLPVKSITENTDGTYKVTVA
VDQLVEEGTDGYKDDYTFTVAKSKAEQPGVYTSFKQLVTAMQSNLSGVYTLASDMTADEVSLGDKQTSYLTGAFTGSLIGSDG
TKSYAIYDLKKPLFDTLNGATVRDLDIKTVSADSKENVAALAKAANSANINNVAVEGKISGAKSVAGLVASATNTVIENSSFT

GKLIANHQDSNKNDTGGIVGNITGNSSRVNKVRVDALISTNARNNNQTAGGIVGRLENGALISNSVATGEIRNGQGYSRVGGI
VGSTWQNGRVNNVVSNVDVGDGYVITGDQYAAADVKNASTSVDNRKADRFATKLSKDQIDAKVADYGITVTLDDTGQDLKRNL
REVDYTRLNKAEAERKVAYSNIEKLMPFYNKDLVVHYGNKVATTDKLYTTELLDVVPMKDDEVVTDINNKKNSINKVMLHFKD
NTVEYLDVTFKENFINSQVIEYNVTGKEYIFTPEAFVSDYTAITNNVLSDLQNVTLNSEATKKVLGAANDAALDNLYLDRQFE
EVKANIAEHLRKVLAMDKSINTTGDGVVEYVSEKIKNNKEAFMLGLTYMNRWYDINYGKMNTKDLSTYKFDFNGNNETSTLDT
IVALGNSGLDNLRASNTVGLYANKLASVKGEDSVFDFVEAYRKLFLPNKTNNEWFKENTKAYIVEMKSDIAEVREKQESPTAD
RKYSLGVYDRISAPSWGHKSMLLPLLTLPEESVYISSNMSTLAFGSYERYRDSVDGVILSGDALRTYVRNRVDIAAKRHRDHY
DIWYNLLDSASKEKLFRSVIVYDGFNVKDETGRTYWARLTDKNIGSIKEFFGPVGKWYEYNSSAGAYANGSLTHFVLDRLLDA
YGTSVYTHEMVHNSDSAIYFEGNGRRREGLGAELYALGLLQSVDSVNSHILALNTLYKAEKDDLNRLHTYNPVERFDSDEALQS
YMHGSYDVMYTLDAMEAKAILAQNNDVKKKWFRKIENYYVRDTRHNKDTHAGNKVRPLTDEEVANLTSLNSLIDNDIINRRSY
DDSREYKRNGYYTISMFSPVYAALSNSKGAPGDIMFRKIAYELLAEKGYHKGFLPYVSNQYGAEAFASGSKTFSSWHGRDVAL
VTDDLVFKKVFNGEYSSWADFKKAMFKQRIDKQDNLKPITIQYELGNPNSTKEVTITTAAQMQQLINEAAAKDITNIDRATSH
TPASWVHLLKQKIYNAYLRTTDDFRNSIYK

SeqID 149
MKFNPNQRYTRWSIRRLSVGVASVVVASGFFVLVGQPSSVRADGLNPTPGQVLPEETSGTKEGDLSEKPGDTVLTQAKPEGVT
GNTNSLPTPTERTEVSEETSPSSLDTLFEKDEEAQKNPELTDVLKETVDTADVDGTQASPAETTPEQVKGGVKENTKDSIDVP
AAYLEKAEGKGPFTAGVNQVIPYELFAGDGMLTRLLLKASDNAPWSDNGTAKNPALPPLEGLTKGKYFYEVDLNGNTVGKQGQ
ALIDQLRANGTQTYKATVKVYGNKDGKADLTNLVATKNVDININGLVAKETVQKAVADNVKDSIDVPAAYLEKAKGEGPFTAG
VNHVIPYELFAGDGMLTRLLLKASDKAPWSDNGDAKNPALSPLGENVKTKGQYFYQVALDGNVAGKEKQALIDQFRANGTQTY
SATVNVYGNKDGKPDLDNIVATKKVTININGLISKETVQKAVADNVKDSIDVPAAYLEKAKGEGPFTAGVNHVIPYELFAGDG
MLTRLLLKASDKAPWSDNGDAKNPALSPLGENVKTKGQYFYQLALDGNVAGKEKQALIDQFRANGTQTYSATVNVYGNKDGKP
DLDNIVATKKVTININGLISKETVQKAVADNVKDSIDVPAAYLEKAKGEGPFTAGVNHVIPYELFAGDGMLTRLLLKASDKAP
WSDNGDAKNPALSPLGENVKTKGQYFYQVALDGNVAGKEKQALIDQFRANGTQTYSATVNVYGNKDGKPDLDNIVATKKVTIK
INVKETSDTANGSLSPSNSGSGVTPMNHNHATGTTDSMPADTMTSSTNTMAGENMAASANKMSDTMMSEDKAMLPNTGETQTS
MASIGFLGLALAGLLGGLGLKNKKEEN

SeqID 150
MKSITKKIKATLAGVAALFAVFAPSFVSAQESSTYTVKEGDTLSEIAETHNTTVEKLAENNHIDNIHLIYVDQELVIDGPVAP
VATPAPATYAAPAAQDETVSAPVAETPVVSETVVSTVSGSEAEAKEWIAQKESGGSYTATNGRYIGRYQLTDSYLNGDYSAEN
QERVADAYVAGRYGSWTAAKNFWLNNGWY

SeqID 151
MNKKKMILTSLASVAILGAGFVTSQPTFVRAEESPQVVEKSSLEKKYEEAKAKADTAKKDYETAKKKAEDAQKKYEDDQKRTE
EKARKEAEASQKLNDVALVVQNAYKEYREVQNQRSKYKSDAEYQKKLTEVDSKIEKARKEQQDLQNKFNEVRAVVVPEPNALA
ETKKKAEEAKAEEKVAKRKYDYATLKVALAKKEVEAKELEIEKLQYEISTLEQEVATAQHQVDNLKKLLAGADPDDGTEVIEA
KLKKGEAELNAKQAELAKKQTELEKLLDSLDPEGKTQDELDKEAEEAELDKKADELQNKVADLEKEISNLEILLGGADPEDDT
AALQNKLAAKKAELAKKQTELEKLLDSLDPEGKTQDELDKEAEEAELDKKADELQNKVADLEKEISNLEILLGGADSEDDTAA
LQNKLATKKAELEKTQKELDAALNELGPDGDEEETPAPAPQPEQPAPAPKPEQPAPAPKPEQPAPAPKPEQPAPAPKPEQPAP
APKPEQPAKPEKPAEEPTQPEKPATPKTGWKQENGMWYFYNTDGSMAIGWLQNNGSWYYLNANGAMATGWVKDGDTWYYLEAS
GAMKASQWFKVSDKWYYVNSNGAMATGWLQYNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWAKVNGSWYYLNANGA
MATGWAKVNGSWYYLNANGSMATGWVKDGDTWYYLEASGAMKASQWFKVSDKWYYVNGLGALAVNTTVDGYKVNANGEWV

SeqID 152
MKKIVLVSLAFLFVLVGCGQKKETGPATKTEKDTLQSALPVIENAEKNTVVTKTLVLPKSDDGSQQTQTITYKDKTFLSLAIQ
QKRPVSDELKTYIDQHGVEETQKALLEAEEKDKSIIEARKLAGFKLETKLLSATELQTTTSFDFQVLDVKKASQLEHLKNIGL
ENLLKNEPSKYISDRLANGATEQ

SeqID 153
MFEVEEWLHSRIGLNFRSGLGRMQQAVDLLGNPEQSYPIIHVTGTNGKGSTIAFMRELFMGHGKKVATFTSPHIVSINDRICI
NGQPIADADFIRLTDQVKEMEKTLLQTPAQLSFFELLTLVAFLYFREQEVDLVLLEVGIGGLLDTTNVVTGEFAVITSIGLDH
QETLGDSLEAIAEQKAGIFKAGKKAVIAKLPPEARLACQKKAESLAVNLYQAGQDFLMLNGDFSSSLLNLSQLNIGLEGVYQQ
ENAALALQTFLLFMRERKEAVDEQAVRKALEQTHWAGRLERIRPQIYLDGAHNLPALTRLAEFIKEKEQEGYRPQILFGSLKR
KDYQGMLGYLTEKLPQVELKVTGFDYQGALDERDVTGYDIVSSYREFISDFEERADAQDLLFVTGSLYFISEVRGYLLDREQI
N

SeqID 154
VGIRVYKPTTNGRRNMTSLDFAEITTSTPEKSLLVALKSKAGRNNNGRITVRHQGGGHKRFYRLVDFKRNKDNVEAVVKTIEY
DPNRSANIALVHYTDGVKAYIIAPKGLEVGQRIVSGPEADIKVGNALPLANIPVGTLIHNIELKPGRGGELVRAAGASAQVLG
SEGKYVLVRLQSGEVRMILGTCRATVGVVGNEQHGLVNLGKAGRSRWKGIRPTVRGSVMNPNDHPHGGGEGKAPVGRKAPSTP
WGKPALGLKTRNKKAKSDKLIVRRRNEK

SeqID 155
MAKKSMVAREAKRQKIVDRYAEKRAALKAAGDYEGLSKLPRNASPTRLHNRCRVTGRPHSVYRKFGLSRIAFRELAHKGQIPG
VTKASW

SeqID 156

MDIRQVTETIAMIEEQNFDIRTITMGISLLDCIDPDINRAAEKIYQKITTKAANLVAVGDEIAAELGIPIVNKRVSVTPISLI
GAATDATDYVVLAKALDKAAKEIGVDFIGGFSALVQKGYQKGDEILINSIPRALAETDKVCSSVNIGSTKSGINMTAVADMGR
IIKETANLSDMGVAKLVVFANAVEDNPFMAGAFHGVGEADVIINVGVSGPGVVKRALEKVRGQSFDVVAETVKKTAFKITRIG
QLVGQMASERLGVEFGIVDLSLAPTPAVGDSVARVLEEMGLETVGTHGTTAALALLNDQVKKGGVMACNQVGGLSGAFIPVSE
DEGMIAAVQNGSLNLEKLEAMTAICSVGLDMIAIPEDTPAETIAAMIADEAAIGVINMKTTAVRIIPKGKEGDMIEFGGLLGT
APVMKVNGASSVDFISRGGQIPAPIHSFKN

SeqID 157
MVNTEVARTTIKTEYFGSLTERMNKYREDVLNKKPYIDAERAVLATRAYERYKEQPNVLKRAYMLKEILENMTIYIEEESMIA
GNQASSNKDAPIFPEYTLEFVLNELDLFEKRDGDVFYITEETKEQLRSIAPFWENNNLRARAGALLPEEVSVYMETGFFGMEG
KMNSGDAHLAVNYQKLLQFGLRGFEERARKAKVALDLTDPASIDKYHFYDSIFIVIDAIKVYAKRFVALAKSLAENANPKRKK
ELLEIADICSRVPYEPATTFAEAIQSVWFIQCILQIESNGHSLSYGRFDQYMYPYMKADLESGKETEDSIVERLTNLWIKTIT
INKVRSQSHTFSSAGSPLYQNVTIGGQTRDKKDAVNPLSYLVLKSVAQTHLPQPNLTVRYHAGLDARFMNECIEVMKLGFGMP
AFNNDEIIIPSFIAKGVLEDDAYDYSAIGCVETAVPGKWGYRCTGMSYMNFPKVLLITMNDGIDPASGKRFAPSFGRFKDMKN
FSELENAWDKTLRYLTRMSVIVENSIDLSLEREVPDILCSALTDDCIGRGKHLKEGGAVYDYISGLQVGIANLSDSLAAIKKL
VFEEERISPSQLWHALETDYAGEEGKVIQEMLIHDAPKYGNDDDYADKLVTAAYDIYVDEIAKYPNTRYGRGPIGGIRYSGTS
SISANVGQGRGTLATPDGRNAGTPLAEGCSPSHNMDQHGPTSVLKSVSKLPTDEIVGGVLLNQKVNPQTLAKEEDKLKLIALL
RTFFNRLHGYHIQYNVVSRETLIDAQKHPEKHRDLIVRVAGYSAFFNVLSKATQDDIIGRTEHTL

SeqID 158
MSQAQYAGTGRRKNAVARVRLVPGTGKITVNKKDVEEYIPHADLRLVINQPFAVTSTVGSYDVFVNVIGGGYAGQSGAIRHGI
ARALLQVDPDFRDSLKRAGLLTRDSRKVERKKPGLKKARKASQFSKR

SeqID 159
LEKKLTIKDIAEMAQTSKTTVSFYLNGKYEKMSQETREKIEKVIHETNYKPSIVARSLNSKRTKLIGVLIGDITNSFSNQIVK
GIEDIASQNGYQVMIGNSNYSQESEDRYIESMLLLGVDGFIIQPTSNFRKYSRIIDEKKKKMVFFDSQLYEHRTSWVKTNNYD
AVYDMTQSCIEKGYEHFLLITADTSRLSTRIERASGFVDALTDANMRHASLTIEDKHTNLEQIKEFLQKEIDPDEKTLVFIPN
CWALPLVFTVIKELNYNLPQVGLIGFDNTEWTCFSSPSVSTLVQPSFEEGQQATKILIDQIEGRNQEERQQVLDCSVNWKEST
F

SeqID 160
MNKGLFEKRCKYSIRKFSLGVASVMIGAAFFGTSPVLADSVQSGSTANLPADLATALATAKENDGRDFEAPKVGEDQGSPEVT
DGPKTEEELLALEKEKPAEEKPKEDKPAAAKPETPKTVTPEWQTVANKEQQGTVTIREEKGVRYNQLSSTAQNDNAGKPALFE
KKGLTVDANGNATVDLTFKDDSEKGKSRFGVFLKFKDTKNNVFVGYDKDGWFWEYKSPTTSTWYRGSRVAAPETGSTNRLSIT
LKSDGQLNASNNDVNLFDTVTLPAAVNDHLKNEKKILLKAGSYDDERTVVSVKTDNQEGVKTEDTPAEKETGPEVDDSKVTYD
TIQSKVLKAVIDQAFPRVKEYSLNGHTLPGQVQQFNQVFINNHRITPEVTYKKINETTAEYLMKLRDDAHLINAEMTVRLQVV
DNQLHFDVTKIVNHNQVTPGQKIDDESKLLSSISFLGNALVSVSSNQTGAKFDGATMSNNTHVSGDDHIDVTNPMKDLAKGYM
YGFVSTDKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNANYVGIHSSEWQWEKAYKGIVFPEYTKELPSAKVVITEDANAD
KNVDWQDGAIAYRSIMNNPQGWEKVKDITAYRIAMNFGSQAQNPFLMTLDGIKKINLHTDGLGQGVLLKGYGSEGHDSGHLNY
ADIGKRIGGVEDFKTLIEKAKKYGAHLGIHVNASETYPESKYFNEKILRKNPDGSYSYGWNWLDQGINIDAAYDLAHGRLARW
EDLKKKLGDGLDFIYVDVWGNGQSGDNGAWATHVLAKEINKQGWRFAIEWGHGGEYDSTFHHWAADLTYGGYTNKGINSAITR
FIRNHQKDAWVGDYRSYGGAANYPLLGGYSMKDFEGWQGRSDYNGYVTNLFAHDVMTKYFQHFTVSKWENGTPVTMTDNGSTY
KWTPEMRVELVDADNNKVVVTRKSNDVNSPQYRERTVTLNGRVIQDGSAYLTPWNWDANGKKLSTDKEKMYYFNTQAGATTWT
LPSDWAKSKVYLYKLTDQGKTEEQELTVKDGKITLDLLANQPYVLYRSKQTNPEMSWSEGMHIYDQGFNSGTLKHWTISGDAS
KAEIVKSQGANDMLRIQGNKEKVSLTQKLTGLKPNTKYAVYVGVDNRSNAKASITVNTGEKEVTTYTNKSLALNYVKAYAHNT
RRDNATVDDTSYFQNMYAFFTTGADVSNVTLTLSREAGDQATYFDEIRTFENNSSMYGDKHDTGKGTFKQDFENVAQGIFPFV
VGGVEGVEDNRTHLSEKHNPYTQRGWNGKKVDDVIEGNWSLKTNGLVSRRNLVYQTIPQNFRFEAGKTYRVTFEYEAGSDNTY
AFVVGKGEFQSGRRGTQASNLEMHELPNTWTDSKKAKKATFLVTGAETGDTWVGIYSTGNASNTRGDSGGNANFRGYNDFMMD
NLQIEEITLTGKMLTENALKNYLPTVAMTNYTKESMDALKEAVFNLSQADDDISVEEARAEIAKIEALKNALVQKKTALVADD
FASLTAPAQAQEGLANAFDGNVSSLWHTSWNGGDVGKPATMVLKEPTEITGLRYVPRGSGSNGNLRDVKLVVTDESGKEHTFT
ATDWPNNNKPKDIDFGKTIKAKKIVLTGTKTYGDGGDKYQSAAELIFTRPQVAETPLDLSGYEAALVKAQKLTDKDNQEEVAS
VQASMKYATDNHLLTERMVEYFADYLNQLKDSATKPDAPTVEKPEFKLRSLASEQGKTPDYKQEIARPETPEQILPATGESQS
DTALILASVSLALSALFVVKTKKD

SeqID 161
MNKPTILRLIKYLSISFLSLVIAAIVLGGGVFFYYVSKAPSLSESKLVATTSSKIYDNKNQLIADLGSERRVNAQANDIPTDL
VKAIVSIEDHRFFDHRGIDTIRILGAFLRNLQSNSLQGGSTLTQQLIKLTYFSTSTSDQTISRKAQEAWLAIQLEQKATKQEI
LTYYINKVYMSNGNYGMQTAAQNYYGKDLNNLSLPQLALLAGMPQAPNQYDPYSHPEAAQDRRNLVLSEMKNQGYISAEQYEK
AVNTPITDGLQSLKSASNYPAYMDNYLKEVINQVEEETGYNLLTTGMDVYTNVDQEAQKHLWDIYNTDEYVAYPDDELQVAST
IVDVSNGKVIAQLGARHQSSNVSFGINQAVETNRDWGSTMKPITDYAPALEYGVYDSTATIVHDEPYNYPGTNTPVYNWDRGY
FGNITLQYALQQSRNVPAVETLNKVGLNRAKTFLNGLGIDYPSIHYSNAISSNTTESDKKYGASSEKMAAAYAAFANGGTYYK
PMYIHKVVFSDGSEKEFSNVGTRAMKETTAYMMTDMMKTVLTYGTGRNAYLAWLPQAGKTGTSNYTDEEIENHIKTSQFVAPD
ELFAGYTRKYSMAVWTGYSNRLTPLVGNGLTVAAKVYRSMMTYLSEGSNPEDWNIPEGLYRNGEFVFKNGARSTWNSPAPQQP
PSTESSSSSSDSSTSQSSSTTPSTNNSTTTNPNNNTQQSNTTPDQQNQNPQPAQP

SeqID 162
MSKKRRNRHKKEGQEPQFDFDEAKELTVGQAIRKNEEVESGVLPEDSILDKYVKQHRDEIEADKFATRQYKKEEFVETQSLDD
LIQEMREAVEKSEASSEEVPSSEDILLPLPLDDEEQGLDPLLLDDENPTEMTEEVEEEQNLSRLDQEDSEKKSKKGFILTVLA

LVSVIICVSAYYVYRQVARSTKEIETSQSTTANQSDVDDFNTLYDAFYTDSNKTALKNSQFDKLSQLKTLLDKLEGSREHTLA
KSKYDSLATQIKAIQDVNAQFEKPAIVDGVLDTNAKAKSDAKFTDIKTGNTELDKVLDKAISLGKSQQTSTSSSSSSQTSSSS
SSQASSNTTSEPKPSSSNETRSSRSEVNMGLSSAGVAVQRSASRVAYNQSAIDDSNNSAWDFADGVLEQILATSRSRGYITGD
QYILERVNIVNGNGYYNLYKPDGTYLFTLNCKTGYFVGNGAGHADDLDY

SeqID 163

MKLLKKMMQVALAVFFFGLLATNTVFANTTGGRFVDKDNRKYYVKDDHKAIYWHKIDGKTYYFGDIGEMVVGWQYLEIPGTGY
RDNLFDNQPVNEIGLQEKWYYFGQDGALLEQTDKQVLEAKTSENTGKVYGEQYPLSAEKRTYYFDNNYAVKTGWIYEEGHWYY
LNKLGNFGDDSYNPLPIGEVAKGWTQDFHVTIDIDRSKPAPWYYLDASGKMLTDWQKVNGKWYYFGSSGSMATGWKYVRGKWY
YLDNKNGDMKTGWQYLGNKWYYLRSSGAMVTGWYQDGSTWYYLDPSNGDMKIGWTKVNGKWYYLNSNGAMVTGSQTIDGKVYN
FASSGEWI

SeqID 164

MKILKKTMQVGLTVFFFGLLGTSTVFADDSEGWQFVQENGRTYYKKGDLKETYWRVIDGKYYYFDSLSGEMVVGWQYIPFPSK
GSTIGPYPNGIRLEGFPKSEWYYFDKNGVLQEFVGWKTLEIKTKDSVGRKYGEKREDSEDKEEKRYYTNYYFNQNHSLETGWL
YDQSNWYYLAKTEINGENYLGGERRAGWINDDSTWYYLDPTTGIMQTGWQYLGNKWYYLRSSGAMATGWYQEGTTWYYLDHPN
GDMKTGWQNLGNKWYYLRSSGAMATGWYQDGSTWYYLNAGNGDMKTGWFQVNGNWYYAYSSGALAVNTTVDGYSVNYNGEWVR

SeqID 165

MVLSKYYGVADGMNVEGRGSANFIKDNVLITAAHNYYRHDYGKEADDIYVLPAVSPSQEPFGKIKVKEVRYLKEFRNLNSKDA
REYDLALLILEEPIGAKLGTLGLPTSQKNLTGITVTITGYPSYNFKIHQMYTDKKQVLSDDGMFLDYQVDTLEGSSGSTVYDA
SHRVVGVHTLGDGANQINSAVKLNERNLPFIYSVLKGYSLEGWKKINGSWYHYRQHDKQTGWQEINDTWYYLDSSGKMLTDWQ
KVNGKWYYLNSNGAMVTGSQTIDGKVYNFASSGEWI

SeqID 166

LMKKTFFLLVLGLFCLLPLSVFAIDFKINSYQGDLYIHADNTAEFRQKIVYQFEEDFKGQIVGLGRAGKMPSGFDIDPHPKIQ
AAKNGAELADVTSEVTEEADGYTVRVYNPGQEGDIVEVDLVWNLKNLLFLYDDIAELNWQPLTDSSESIEKFEFHVRGDKGAE
KLFFHTGKLFREGTIEKSNLDYTIRLDNLPAKRGVELHAYWPRTDFASARDQGLKGNRLEEFNKIEDSIVREKDQSKQLVTWV
LPSILSISLLLSVCFYFIYRRKTTPSVKYAKNHRLYEPPMELEPMVLSEAVYSTSLEEVSPLVKGAGKFTFDQLIQATLLDVI
DRGNVSIISEGDAVGLRLVKEDGLSSFEKDCLNLAFSGKKEETLSNLFADYKVSDSLYRRAKVSDEKRIQARGLQLKSSFEEV
LNQMQEGVRKRVSFWGLPDYYRPLTGGEKALQVGMGALTILPLFIGFGLFLYSLDVHGYLYLPLPILGFLGLVLSVFYYWKLR
LDNRDGVLNEAGAEVYYLWTSFENMLREIARLDQAELESIVVWNRLLVYATLFGYADKVSHLMKVHQIQVENPDINLYVAYGW
HSTFYHSTAQMSHYASVANTASTYSVSSGSGSSGGGFSGGGGGGSIGAF

SeqID 167

MKSINKFLTMLAALLLTASSLFSAATVFAAGTTTTSVTVHKLLATDGDMDKIANELETGNYAGNKVGVLPANAKEIAGVMFVW
TNTNNEIIDENGQTLGVNIDPQTFKLSGAMPATAMKKLTEAEGAKFNTANLPAAKYKIYEIHSLSTYVGEDGATLTGSKAVPI
EIELPLNDVVDAHVYPKNTEAKPKIDKDFKGKANPDTPRVDKDTPVNHQVGDVVEYEIVTKIPALANYATANWSDRMTEGLAF
NKGTVKVTVDDVALEAGDYALTEVATGFDLKLTDAGLAKVNDQNAEKTVKITYSATLNDKAIVEVPESNDVTFNYGNNPDHGN
TPKPNKPNENGDLTLTKTWVDATGAPIPAGAEATFDLVNAQTGKVVQTVTLTTDKNTVTVNGLDKNTEYKFVERSIKGYSADY
QEITTAGEIAVKNWKDENPKPLDPTEPKVVTYGKKFVKVNDKDNRLAGAEFVIANADNAGQYLARKADKVSQEEKQLVVTTKD
ALDRAVAAYNALTAQQQTQQEKEKVDKAQAAYNAAVIAANNAFEWVADKDNENVVKLVSDAQGRFEITGLLAGTYYLEETKQP
AGYALLTSRQKFEVTATSYSATGQGIEYTAGSGKDDATKVVNKKITIPQTGGIGTIIFAVAGAAIMGIAVYAYVKNNKDEDQL
A

SeqID 168

MAVMAYPLVSRLYYRVESNQQIADFDKEKATLDEADIDERMKLAQAFNDSLNNVVSGDPWSEEMKKKGRAEYARMLEIHERMG
HVEIPVIDVDLPVYAGTAEEVLQQGAGHLEGTSLPIGGNSTHAVITAHTGLPTAKMFTDLTKLKVGDKFYVHNIKEVMAYQVD
QVKVIEPTNFDDLLIVPGHDYVTLLTCTPYMINTHRLLVRGHRIPYVAEVEEEFIAANKLSHLYRYLFYVAVGLIVILLWIIR
RLRKKKKQPEKALKALKAARKEVKVEDGQQ

SeqID 169

MSRTKLRALLGYLLMLVACLIPIYCFGQMVLQSLGQVKGHATFVKSMTTEMYQEQQNHSLAYNQRLASQNRIVDPFLAEGYEV
NYQVSDDPDAVYGYLSIPSLEIMEPVYLGADYHHLGMGLAHVDGTPLPLDGTGIRSVIAGHRAEPSHVFFRHLDQLKVGDALY
YDNGQEIVEYQMMDTEIILPSEWEKLESVSSKNIMTLITCDPIPTFNKRLLVNFERVAVYQKSDPQTAAVARVAFTKEGQSVS
RVATSQWLYRGLVVLAFLGILFVLWKLARLLRGK

SeqID 170

MKNPFFERRCRYSIRKLSVGACSLMIGAVLFAGPALAEETAVPENSGANTELVSGESEHSTNEADKQNEGEHARENKLEKAEG
VAIASETASPASNEAATTETAEAASAAKPEEKASEVVAETPSAEAKPKSDKETEAKPEATNQGDESKPAAEANKTEKEVQPDV
PKNTEKTLKPKEIKFNSWEELLKWEPGAREDDAINRGSVVLASRRTGHLVNEKASKEAKVQALSNTNSKAKDHASVGGEEFKA
YAFDYWQYLDSMVFWEGLVPTPDVIDAGHRNGVPVYGTLFFNWSNSIADQERFAEALKQDADGSFPIARKLVDMAKYYGYDGY
FINQETTGDLVKPLGEKMRQFMLYSKEYAAKVNHPIKYSWYDAMTYNYGRYHQDGLGEYNYQFMQPEGDKVPADNFFANFNWD
KAKNDYTIATANWIGRNPYDVFAGLELQQGGSYKTKVKWNDILDENGKLRLSLGLFAPDTITSLGKTGEDYHKNEDIFFTGYQ
GDPTGQKPGDKDWYGIANLVADRTPAVGNTFTTSFNTGHGKKWFVDGKVSKDSEWNYRSVSGVLPTWRWWQTSTGEKLRAEYD
FTDAYNGGNSLKFSGDVAGKTDQDVRLYSTKLEVTEKTKLRVAHKGGKGSKVYMAFSTTPDYKFDDADAWKELTLSDNWTNEE
FDLSSLAGKTIYAVKLFFEHEGAVKDYQFNLGQLTISDNHQEPQSPTSFSVVKQSLKNAQEAEAVVQFKGNKDADFYEVYEKD
GDSWKLLTGSSSTTIYLPKVSRSASAQGTTQELKVVAVGKNGVRSEAATTTFDWGMTVKDTSLPKPLAENIVPGATVIDSTFP

KTEGGEGIEGMLNGTITSLSDKWSSAQLSGSVDIRLTKPRTVVRWVMDHAGAGGESVNDGLMNTKDFDLYYKDADGEWKLAKE
VRGNKAHVTDITLDKPITAQDWRLNVVTSDNGTPWKAIRIYNWKMYEKLDTESVNIPMAKAAARSLGNNKVQVGFADVPAGAT
ITVYDNPNSQTPLATLKSEVGGDLASAPLDLTNQSGLLYYRTQLPGKEISNVLAVSVPKDDRRIKSVSLETGPKKTSYAEGED
LDLRGGVLRVQYEGGTEDELIRLTHAGVSVSGFDTHHKGEQNLTLQYLGQPVNANLSVTVTGQDEASPKTILGIEVSQEPKKD
YLVGDSLDLSEGRFAVAYSNDTMEEHSFTDEGVEISGYDAQKTGRQTLTLHYQGHEVSFDVLVSPKAALNDEYLKQKLAEVEA
AKNKVVYNFASSEVKEAFLKAIEAAEQVLKDHETSTQDQVNDRLNKLTEAHKALNGQEKFTEEKTELDRLTGEVQELLAAKPN
HPSGSALAPLLEKNKALVKVDLSPEELTTAKQSLKDLVALLKEDKPAVFSDSKTGVEVHFSNKEKTVIKGLKVERVQASAEE
KKYFAGEDAHVFEIEGLDEKGQDVDLSYASIVKIPIEKDKKVKKVFFLPEGKEAVELAFEQTDSHVIFTAPHFTHYAFVYESA
EKPQPAKPAPQNTVLPKPTYQPTSDQQKAPKLEVQEEKVAFHRQEHENTEMLVGEQRVIIQGRDGLLRHVFEVDENGQRRLRS
TEVIQEAIPEIVEIGTKVKTVPAVVATQEKPAQNTAVKSEEASKQLPNTGTADANEALIAGLASLGLASLALTLRRKREDKD

SeqID 171
MSITSFVKRIQDITRNDAGVNGDAQRIEQMSWLLFLKIYDSREMVWELEEDEYESIIPEELKWRNWAHAQNGERVLTGDELLD
FVNNKLFKELKELEITSNMPIRKTIVKSAFEDANNYMKNGVLLRQVINVIDEVDFNSPEDRHSFNDIYEKILKDIQNAGNSGE
FYTPRAATDFIAEVLDPKLGESMADLACGTGGFLTSTLNRLSSQRKTSEDTKKYNTAVFGIEKKAFPHLLAVTNLFLHEIDDP
KIVHGNTLEKNVREYTDDEKFDIIMMNPPFGGSELETIKNNFPAELRSSETADLFMAVIMYRLKENGRVGVILPDGFLFGEGV
KTRLKQKLVDEFNLHTIIRLPHSVFAPYTGIHTNILFFDKTKKTEETWFYRLDMPDGYKNFSKTKPMKSEHFNPVRDWWENRE
EILEGKFYKSKSFTPSELAELNYNLDQCDFPKEEEEILNPFELIQNYQAERATLNHKIDNVLADILQLLEDK

SeqID 172
MNNTEFYDRLGVSKNASADEIKKAYRKLSKKYHPDINKEPGAEDKYKEVQEAYETLSDDQKRAAYDQYGAAGANGGFGGAGGF
GGFNGAGGFGGFEDIFSSFFGGGGSSRNPNAPRQGDDLQYRVNLTFEEAIFGTEKEVKYHREAGCRTCNGSGAKPGTSPVTCG
RCHGAGVINVDTQTPLGMMRRQVTCDVCHGRGKEIKYPCTTCHGTGHEKQAHSVHVKIPAGVETGQQIRLAGQGEAGFNGGPY
GDLYVVVSVEASDKFEREGTTIFYNLNLNFVQAALGDTVDIPTVHGDVELVIPEGTQTGKKFRLRSKGAPSLRGGAVGDQYVT
VNVVTPTGLNDRQKVALKEFAAAGDLKVNPKKKGFFDHIKDAFDGE

SeqID 173
MNPNLFRSVEFYQRRYHNYATVLIIPLSLLFTFILIFSLVATKEITVTSQGEIAPTSVIASIQSTSDNPILANHLVANQVVEK
GDLLIKYSETMEESQKTALATQLQRLEKQKEGLGILKQSLEKATDLFSGEDEFGYHNTFMNFTKQSHDIELGITKTNTEVSNQ
ANLSNSSSSAIEQEITKVQQQIGEYQELRDAIINNRARLPTGNPHQSILNRYLVASQGQTQGTAEEPFLSQINQSIAGLESSI
ASLKIQQAGIGSVATYDNSLATKIEVLRTQFLQTASQQQLTVENQLTELKVQLDQATQRLENNTLTSPSKGIVHLNSEFEGKN
RIPTGTEIAQIFPVITDTREVLITYYVSSDYLPLLDKGQTVRLKLEKIGNHGTTIIGQLQTIDQTPTRTEQGNLFKLTALAKL
SNEDSKLIQYGLQGRVTSVTTKKTYFDYFKDKILTHSD

SeqID 174
MSKKLNRKKQLRNGLRRAGAFSSTVTKVVDETKKVVKRAEQSASAAGKAVSKKVEQAVEATKEQAQKVANSVEDFAANLGGLP
LDRAKTFYDEGIKSASDFKNWTEKELLALKGIGPATIKKLKENGIKFK

SeqID 175
LISLFGLAAAKPVQADTSIADIQKRGELVVGVKQDVPNFGYKDPKTGTYSGIETDLAKMVADELKVKIRYVPVTAQTRGPLLD
NEQVDMDIATFTITDERKKLYNFTSPYYTDASGFLVNKSAKIKKIEDLNGKTIGVAQGSITQRLITELGKKKGLKFKFVELGS
YPELITSLHAHRIDTFSVDRSILSGYTSKRTALLDDSFKPSDYGIVTKKSNTELNDYLDNLVTKWSKDGSLQKLYDRYKLKPS
SHTAD

SeqID 176
MSNISLTTLGGVRENGKNMYIAEIGESIFVLNVGLKYPENEQLGVDVVIPNMDYLFENSDRIAGVFLTHGHADAIGALPYLLA
EAKVPVFGSELTIELAKLFVKGNDAVKKFNDFHVIDENTEIDFGGTVVSFFPTTYSVPESLGIVLKTSEGSIVYTGDFKFDQT
ASESYATDFARLAEIGRDGVLALLSDSANADSNIQVASESEVRDEITQTIADWEGRIIVAAVSSNLSRIQQIFDAADKTGRRI
VLTGFDIENIVRTAIRLKKLSLANEILLIKPKDMSRFEDHELIILETGRMGEPINGLRKMSIGRHRYVEIKDGDLVYIATAPS
IAKEAFVARVENMIYQAGGVVKLITQSLHVSGHGNVRDLQLMINLLQPKYLFPVQGEYRELDAHAKAAMAVGMLPERIFIPKK
GTTMAYENGDFVPAGSVSAGDILIDGNAIGDVGNVVLRDRKVLSEDGIFIVAITVNRREKKIVARARVHTRGFVYLKKSRDIL
RESSELINQTVEEYLQGDDFDWADLKGKVRDNLTKYLFDQTKRRPAILPVVMEAK

SeqID 177
MKKSTVLSLTTAAVILAAYAPNEVVLADTSSSEDALNISDKEKVAENKEKHENIHSAMETSQDFKEKKTAVIKEKEVVSKNPV
IDNNTSNEEAKIKEENSNKSQGDYTDSFVNKNTENPKKEDKVVYIAEFKDKESGEKAIKELSSLKNTKVLYTYDRIFNGSAIE
TTPDNLDKIKQIEGISSVERAQKVQPMMNHARKEIGVEEAIDYLKSINAPFGKNFDGRGMVISNIDTGTDYRHKAMRIDDDAK
ASMRFKKEDLKGTDKNYWLSDKIPHAFNYYNGGKITVEKYDDGRDYFDPHGMHIAGILAGNDTEQDIKNFNGIDGIAPNAQIF
SYKMYSDAGSGFAGDETMFHAIEDSIKHNVDVVSVSSGFTGTGLVGEKYWQAIRALRKAGIPMVVATGNYATSASSSSWDLVA
NNHLKMTDTGNVTRTAAHEDAIAVASAKNQTVEFDKVNIGGESFKYRNIGAFFDKSKITTNEDGTKAPSKLKFVYIGKGQDQD
LIGLDLRGKIAVMDRIYTKDLKNAFKKAMDKGARAIMVVNTVNYYNRDNWTELPAMGYEADEGTKSQVFSISGDDGVKLWNMI
NPDKKTEVKRNNKEDFKDKLEQYYPIDMESFNSNKPNVGDEKEIDFKFAPDTDKELYKEDIIVPAGSTSWGPRIDLLLKPDVS
APGKNIKSTLNVINGKSTYGYMSGTSMATPIVAASTVLIRPKLKEMLERPVLKNLKGDDKIDLTSLTKIALQNTARPMMDATS
WKEKSQYFASPRQQGAGLINVANALRNEVVATFKNTDSKGLVNSYGSISLKEIKGDKKYFTIKLHNTSNRPLTFKVSASAITT
DSLTDRLKLDETYKDEKSPDGKQIVPEIHPEKVKGANITFEHDTFTIGANSSFDLNAVINVGEAKNKNKFVESFIHFESVEEM
EALNSNGKKINFQPSLSMPLMGFAGNWNHEPILDKWAWEEGSRSKTLGGYDDDGKPKIPGTLNKIGGEHGIDKFNPAGVIQN
RKDKNTTSLDQNPELFAFNNEGINAPSSSGSKIANIYPLDSNGNPQDAQLERGLTPSPLVLRSAEEGLISIVNTNKEGENQRD
LKVISREHFIRGILNSKSNDAKGIKSSKLKVWGDLKWDGLIYNPRGREENAPESKDNQDPATKIRGQFEPIAEGQYFYKFKYR

LTKDYPWQVSYIPVKIDNTAPKIVSVDFSNPEKIKLITKDTYHKVKDQYKNETLFARDQKEHPEKFDEIANEVWYAGAALVNE
DGEVEKNLEVTYAGEGQGRNRKLDKDGNTIYEIKGAGDLRGKIIEVIALDGSSNFTKIHRIKFANQADEKGMISYYLVDPDQD
SSKYQKLGEIAESKFKNLGNGKEGSLKKDTTGVEHHHQENEESIKEKSSFTIDRNISTIRDFENKDLKKLIKKKFREVDDFTS
ETGKRMEEYDYKYDDKGNIIAYDDGTDLEYETEKLDEIKSKIYGVLSPSKDGHFEILGKISNVSKNAKVVYYGNNYKSIEIKAT
KYDFHSKTMTFDLYANINDIVDGLAFAGDMRLFVKDNDQKKAEIKIRMPEKIKETKSEYPYVSSYGNVIELGEGDLSKNKPDN
LTKMESGKIYSDSEKQQYLLKDNIILRKGYALKVTTYNPGKTDMLEGNGVYSKEDIAKIQKANPNLRALSETTIYADSRNVED
GRSTQSVLMSALDGFNIIRYQVFTFKMNDKGEAIDKDGNLVTDSSKLVLFGKDDKEYTGEDKFNVEAIKEDGSMLFIDTKPVN
LSMDKNYFNPSKSNKIYVRNPEFYLRGKISDKGGFNWELRVNESVVDNYLIYGDLHIDNTRDFNIKLNVKDGDIMDWGMKDYK
ANGFPDKVTDMDGNVYLQTGYSDLNAKAVGVHYQFLYDNVKPEVNIDPKGNTSIEYADGKSVVFNINDKRNNGFDGEIQEQHI
YINGKEYTSFNDIKQIIDKTLNIKIVVKDFARNTTVKEFILNKDTGEVSELKPHRVTVTIQNGKEMSSTIVSEEDFILPVYKG
ELEKGYQFDGWEISGFEGKKDAGYVINLSKDTPIKPVFKKIEEKKEEENKPTFDVSKKKDNPQVNHSQLNESHRKEDLQREEH
SQKSDSTKDVTATVLDKNNISSKSTTNNPNKLPKTGTASGAQTLLAAGIMPIVGIPLGLKKKNQD

SeqID 178
MGKGHWNRKRVYSIRKFAVGACSVMIGTCAVLLGGNIAGESVVYADETLITHTAEKPKEEKMIVEEKADKALETKNIVERTEC
SEPSSTEAIASEKKEDEAVTPKEEKVSAKPEEKAPRIESQASNQEKPLKEDAKAVTNEEVNQMIEDRKVDFNQWYFKLNANS
KEAIKPDADVSTWKKLDLPYDWSIFNDFDHESPAQNEGGQLNGGEAWYRKTFKLDEKDLKKNVRLTFDGVYMDSQVYVNGQLV
GHYPNGYNQFSYDITKYLQKDGRENVIAVHAVNKQPSSRWYSGSGIYRDVTLQVTDKVHVEKNGTTILTPKLEEQQHGKVETE
VTSKIVNTDDKDHELVAEYQIVERGGHAVTGLVRTASRTLKAHESTSLDAILEVERPKLWTVLNDKPALYELITRVYRDGQLV
DAKKDLFGYRYYHWTPNEGFSLNGERIKFHGVSLHHDHGALGAEENYKAEYRRLKQMKEMGVNSIRTTHNPASEQTLQIAAEI
GLLVQEEAFDTWYGGKKPYDYGRFFEKDATHPEARKGEKWSDFDLRTMVERGKNNPAIFMWSIGNEIGEANGDAHSLATVKRI
VKVIKDVDKTRYVTMGADKFRFGNGSGGHEKIADELDAVGFNYSEDNYKALRAKHPKWLIYGSETSSATRTRGSYYRPERELK
HSNGPERNYEQSDYGNDRVGWGKTATASWTFDRDNAGYAGQFIWTGTDYIGEPTPWHNQNQTPVKSSYFGIVDTAGIPKHDFY
LYQSQWVSVKKKPMVHLLPHWNWENKELASKVADSEGKIPVRAYSNASSVELFLNGKSLGLKTFNKKQTSDGRTYQEGANANE
LYLEWKVAYQPGTLEAIARDESGKEIARDKITTAGKPAAVRLIKEDHAIAADGKDLTYIYYEIVDSQGNVVPTANNLVRFQLK
GQGQLVGVDNGEQASRERYKAQADGSWIRKAFNGKGVAIVKSTEQAGKFTLTAHSDLLKSNQVTVFTGKKEGQEKTVLGTEVI
KVQTIIGEAPEMPTTVPFVYSDGSRAERPVTWSSVDVSKPGIVTVKGMADGREVEARVEVIALKSELPVVKRIAPNTDLNSVI
KSVSYVLIDGSVEEYEVDKWEIAEEDKAKLAIPGSRIQATGYLEGQPIHATLVVEEGNPAAPAVPTVTVGGEAVTGLTSQKPM
QYRTLAYGAKLPEVTASAKNAAVTVLQASAANGMRASIFIQPKDGGPLQTYAIQFLEEAPKIAHLSLQVEKADSLKEDQTVKI
SVRAHYQDGTQAVLPADKVTFSTSGEGEVAIRKGMLELHKPGAVTLNAEYEGAKDQVELTIQANTEKKIAQSIRPVNVVTDLK
QEPSLPATVTVEYDKGFPKTHKVTWQAIPKEKLDSYQTFEVLGKVEGIDLEARAKVSVEGIVSVEEVSVTTPIAEAPQLPESV
RTYDSNGHVSSAKVAWDAIRPEQYAKEGVFTVNGRLEGTQLTTKLHVRVSAQTEQGANISDQWTGSELPLAFASDSNPSDPVS
NVNDKLISYNNQPANRWTNWNRTNPEASVGVLFGDSGILSKRSVDNLSVGFHEDHGVGVPKSYVIEYVVGKTVPTAPKNPSFV
GNEDHVFNDSANWKPVTNLKAPAQLKAGEMNHFSFDKVETYAVRIRMVKADNKRGTSITEVQIFAKQVAAAKQGQTRIQVDGI
DLANFNPDLTDYYLESVDGKVPAVTASVSNNGLATVVPSVREGEPVRVIAKAENGDILGEYRLHFTKDKSLLSHKPVAAVKQA
RLLQVGQALELPTKVPVYFTGKDGYETKDLTVEWEEVPAENLTKAGQFTVRGRVLGSNLVAEITVRVTDKLGETLSDNPNYDK
NSNQAFASATNDIDKNSHDRVDYLNDGDHSENRRWTNWSPTPSSNPEVSAGVIFRENGKIVERTVTQGKVQFFADSGTDAPSK
LVLERYVGPEFEVPTYYSNYQAYDADHPFNNPENWEAVPYRADKDIAAGDEINVTFKAIKAKAMRWRMERKADKSGVAMIEMT
FLAPSELPQESTQSKILVDGKELADFAENRQDYQITYKGQRPKVSVEENNQVASTVVDSGEDSFPVLVRLVSESGKQVKEYRT
HLTKEKPVSEKTVAAVQEDLPKIEFVEKDLAYKTVEKKDSTLYLGETRVEQEGKVGKERIFTAINPDGSKEEKLREVVVEVPTI
RIVLVGTKPVAQEAKKPQVSEKADTKPIDSSEASQTNKAQLPSTGSAASQAAVAAGLTLLGLSAGLVVTKGKKED

SeqID 179
MAPSVVDAATYHYVNKEIISQEAKDLIQTGKPDRNEVVYGLVYQKDQLPQTGTEASVLTAFGLLTVGSLLLIYKRKKIASVFI
VGAMGLVVLPSAGAVDPVATLALASREGVVEMEGYRYVGYLSGDILKTLGLDTVLEETSAKPGEVTVVEVETPQSITNQEQAI
TENQVVETEEAPKEEAPKTEESPKEEPKSEVKPTDDTLPKVEEGKEDSAEPAPVEEVGGEVESKPEEKVAVKPESQPSDKPAI
ESKVEQAGEPVAPREDEKAPVEPEKQPEAPEEEKAVEETPKQEESTPDTKAEETVEPKEETVNQSIEQPKVETPAVEKQTEPT
EEPKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEEEKAVEETPKPEDKIKGIGTKEPVDKSELNNQIDKASSVSPTDYSTAS
YNALGPVLETAKGVYASEPVKQPEVNSETNKLKTAIDALNVDKTELNNTIADAKTKVKEHYSDRSWQNLQTEVTKAEKVAANT
DAKQSEVNEAVEKLTATIEKLVELSEKPILTLTSTDKKILEREAVAKYTLENQNKTKIKSITAELKKGEEVINTVVLTDDKVT
TETISAAFKNLEYYKEYTLSTTMIYDRGNGEETETLENQNIQLDLKKVELKNIKRTDLIKYENGKETNESLITTIPDDKSNYY
LKITSNNQKTTLLAVKNIEETTVNGTPVYKVTAIADNLVSRTADNKFEEEYVHYIEKPKVHEDNVYYNFKELVEAIQNDPSKI
YRLGQSMSARNVVPNGKSYITKEFTGKLLSSEGKQFAITELEHPLFNVITNATINNVNFENVEIERSGQDNIASLANTMKGSS
VITNVKITGTLSGRNNVAGFVNNMNDGTRIENVAFFGKLHSTSGNGSHTGGIAGTNYRGIVRKAYVDATITGNKTRASLLVPK
VDYGLTLDHLIGTKALLTESVVKGKIDVSNPVEVGAIASKTWPVGTVSNSVSYAKIIRGEELFGSNDVDDSDYASAHIKDLYA
VEGYSSGNRSFRKSKTFTKLTKEQADAKVTTFNITADKLESDLSPLAKLNEEKAYSSIQDYNAEYNQAYKNLEKLIPFYNKDY
IVYQGNKLNKEHHLNTKEVLSVTAMNNNEFITNLDEANKIIVHYADGTKDYFNLSSSSEGLSNVKEYTITDLGIKYTPNIVQL
DNTTLVNDIKSILESVELQSQTMYQHLNRLGDYRVNAIKDLYLEESFTDVKENLTNLITKLVQNEEHQLNDSPAARQMIRDKV
EKNKAALLLGLTYLNRYYGVKFGDVNIKELMLFKPDFYGEKVSVLDRLIEIGSKENNIKGSRTFDAFGQVLAKYTKSGNLDAI
LNYNRQLFTNIDNMNDWFIDATEDHVYIAERASEVEEIKNSKHRAFDNLKRSHLRNTILPLLNIDKAHLYLISNYNAIAFGSA
ERLGKKSLEDIKDIVNKAADGYRNYYDFWYRLASDNVKQRLLRDAVIPIWEGYNAPGGWVEKYGRYNTDKVYTPLREFFGPMI
KYYNYNGTGAYAAIYPNSDDIRTDVKYVHLEMVGEYGISVYTHETTHVNDRAIYLGGFGHREGTDAEAYAQGMLQTPVTGSGI
DEFGSLGINMVFKRKNDGNQWYITDPKTLKTREDINRYMKGYNDTLTLLDEIEAESVISQQNKDLNSAWFKKIDREYRDNNKI
NQWDKIRNLSQEEKNELNIQSVNDLVDQQLMTNRNPGNGIYKPEAISYNDQSPYVGVRMMTGIYGGNTSKGAPGAVSFKHNAI
RLWGYYGYENGFLGYASNKYKQQSKTDGESVLSDEYIIKKISNNTFNTIEEFKKAYFKEVKDKATKGLTTFEVNGSSVSSYDI
LLTLFKEAVKKDAETLKQEANGNKTVSMNNTVKLKEAVYKKLLQQTNSFKTSIFK

SeqID 180

MNKRLFSKMSLVTLPILALFSQSVLAEENIHFSSCKEAWANGYSDIHEGEPGYSAKLDRDHDGVACELKNAPKGAFKAKQSTA
IQINTSSATTSGWVKQDGAWYYFDGNGNLVKNAWQGSYYLKADGKMAQSEWIYDSSYQAWYYLKSDGSYAKNAWQGAYYLKSN
GKMAQGEWVYDSSYQAWYYLKSDGSYARNAWQGNYYLKSDGKMAKGEWVYDATYQAWYYLTSDGSYAYSTWQGNYYLKSDGKM
AVNEWVDGGRYYVGADGVWKEVQASTASSSNDSNSEYSAALGKAKSYNSLFHMSKKRMYRQLTSDFDKFSNDAAQYAIDHLDD

SeqID 181
MKVIDQFKNKKVLVLGLAKSGESAARLLDKLGAIVTVNDGKPFEDNPAAQSLLEEGIKVITGGHPLELLDEEFALMVKNPGIP
YNNPMIEKALAKGIPVLTEVELAYLISEAPIIGITGSNGKTTTTTMIGEVLTAAGQHGLLSGNIGYPASQVAQIASDKDTLVM
ELSSFQLMGVQEFHPEIAVITNLMPTHIDYHGSFSEYVAAKWNIQNKMTAADFLVLNFNQDLAKDLTSKTEATVVPFSTLEKV
DGAYLEDGQLYFRGEVVMAANEIGVPGSHNVENALATIAVAKLRDVDNQTIKETLSAFGGVKHRLQFVDDIKGVKFYNDSKST
NILATQKALSGFDNSKVVLIAGGLDRGNEFDELVPDITGLKKMVILGQSAERVKRAADKAGVAYVEATDIADATRKAYELATQ
GDVVLLSPANASWDMYANFEVRGDLFIDTVAELKE

SeqID 182
MKKKFALSFVALASVALLAACGEVKSGAVNTAGNSVEEKTIKIGPFNFEESGSLAAYGTAEQKGAQLAVDEINAAGGIDGKQIE
VVDKDNKSETAEAASVTTNLVTQSKVSAVVGPATSGATAAAVANATKAGVPLISPSATQDGLTKGQDYLFIGTFQDSFQGKII
SNYVSEKLNAKKVVLYTDNASDYAKGIAKSFRESYKGEIVADETFVAGDTDFQAALTKMKGKDFDAIVVPGYYNEAGKIVNQP
RGMGIDKPIVGGDGFNGEEFVQQATAEKASNIYFISGFSTTVEVSAKAKAFLDAYRAKYNEEPSTFAALAYDSVHLVANAAKG
AKNSGEIKNNLAKTKDFEGVTGQTSFDADHNTVKTAYMMTMNNGKVEAAEVVKP

SeqID 183
MSILEVKNLSHGFGDRAIFEDVSFRLLKGEHIGLVGANGEGKSTFMSIVTGKMLPDEGKVEWSKYVTAGYLDQHSVLAERQSV
RDVLRTAFDELFKAEARINDLYMKMAEDGADVDALMEEVGELQDRLESRDFYTLDAKIDEVARALGVMDFGMDTDVTSLSGGQ
RTKVLLAKLLLEKPDILLLDEPTNYLDAEHIDWLKRYLQNYENAFVLISHDIPFLNDVINIVYHVENQQLTRYSGDYYQFQEV
YAMKKSQLEAAYERQQKEIADLKDFVARNKARVATRNMAMSRQKKLDKMDIIELQSEKPKPSFDFKPARTPGRFIFQAKNLQI
GYDRPLTKPLNLTFERNQKVAIIGANGIGKTTLLKSLLGIISPIAGEVERGDYLELGYFEQEVEGGNRQTPLEAVWNAFPALN
QAEVRAALARCGLTTKHIESQIQVLSGGEQAKVRFCLLMNRENNVLVLDEPTNHLDVDAKDELKRALKEYRGSILMVCHEPDE
YEGWIDQIWDFNNLT

SeqID 184
MKKKNGKAKKWQLYAAIGAASVVVLGAGGILLFRQPSQTALKDEPTHLVVAKEGSVASSVLLSGTVTAKNEQYVYFDASKGDI
DEILVSVGDKVSEGQALVKYSSSEAQAAYDSASRAVARADRHINELNQARNEAASAPAPQLPAPVGGEDATVQSPTPVAGNSV
ASIDAQLGDARDARADAAAQLSKAQSQLDATTVLSTLEGTVVEVNSNVSKSPTGASQVMVHIVSNENLQVKGELSEYNLANLS
VGQEVSFTSKVYPDKKWTGKLSYISDYPKNNGEAASPAAGNNTGSKYPYTIDVTGEVGDLKQGFSVNIEVKSKTKAILVPVSS
LVMDDSKNYVWIVDEQQKAKKVEVSLGNADAENQEITSGLTNGAKVISNPTSSLEEGKEVKADEATN

SeqID 185
SETNHEIDSNFAGRLNILRAGVLDANDGIISIAGVVIGVASATTNIWIIFLSGFTAILAGAFSMAGGEYVSVSTPKDTEEAAV
SREKLLLDQDRELAKKSLYAAYIQNGEFKTSAQLLTNKIFLKNPLKALVEEKYGIEYEEFTNPWHAAISSFVAFFLRSLPPMI
SVTIFPSDYRIPATVLIVGVALLLTGYTSARLGKAPTKTAMIRNLAIGLLTMGVTFLLGQLFSI

SeqID 186
MKKKLTSLALVGAFLGLSWYGNVQAQESSGNKIHFINVQEGGSDAIILESNGHFAMVDTGEDYDFPDGSDSRYPWREGIETSY
KHVLTDRVFRRLKELGVQKLDFILVTHTHSDHIGNVDELLSTYPVDRVYLKKYSDSRITNSERLWDNLYGYDKVLQTAAEKGV
SVIQNITQGDAHFQFGDMDIQLYNYENETDSSGELKKIWDDNSNSLISVVKVNGKKIYLGGDLDNVHGAEDKYGPLIGKVDLN
KFNHHHDTNKSNTKDFIKNLSPSLIVQTSDSLPWKNGVDSEYVNWLKERGIERINAASKDYDATVFDIRKDGFVNISTSYKPI
PSFQAGWHKSAYGNWWYQAPDSTGEYAVGWNEIEGEWYYFNQTGILLQNQWKKWNNHWFYLTDSGASAKNWKKIAGIWYYFNF
ENQMEIGWIQDKEQWYYLDVDGSMKTGWLQYMGQWYYFAPSGEMKMGWVKDKETWYYMDSTGVMKTGEIEVAGQHYYLEDSGA
MKQGWHKKANDWYFYKTDGSRAVGWIKDKDKWYFLKENGQLLVNGKTPEGYTVDSSGAWLVDVSIEKSATIKTTSHSEIKESF
EVVKKDLENKETSQHESVTNFSTSQDLTSSTSQSSETSVNKSESEQ

SeqID 187
MDLGPTQRGISVVSQSYINVIGAGLAGSEAAYQIAERGIPVKLYEMRGVKSTPQHKTDNFAELVCSNSLRGDALTNAVGLLKF
EMRRLGSVILESAEATRVPAGGALAVDRDGFSQMVTEKVANHPLIEVVRDEITELPTDVITVIATGPLTSDALAEKIHALNDG
AGFYFYDAAAPIIDVNTIDMSKVYLKSRYDKGEAAYLNAPMTKQEFMDFHEALVNAEEAPLSSFEKEKYFEGCMPIEVMAKRG
IKTMLYGPMKPVGLEYPDDYTGPRDGEFKTPYAVVQLRQDNAAGSLYNIVGFQTHLKWGEQKRVFQMIPGLENAEFVRYGVMF
RNSYMDSPNLLEQTYRSKKQPNLFFAGQMTGVEGYVESAASGLVAGINAARLFKEESEAIFPETTAIGSLAHYITHADSKHFQ
PMNVNFGIIKELEGERIRDKKARYEKIAERALADLEEFLTV

SeqID 188
MLIGIPKEIKNNENRVALTPAGVHSLVSRGHRVLIETNAGLGSGFTDADYQKQGAEIVATAGEAWAAELVVKVKESLSSEYGY
LRDDLLLFTYLHMAAAPELADAMLTAKTTETVRDNQGQLPLLVPMSEVAGRMAVQIGAHFLTKQAGGSGVLLGGVPGVPKGKV
TIIGGGVVGTHAARIALGLGAQVTILDISSKRLSVLEEVFGSQIQTLMSNSFNIEASVRDADVVIGAILIPGAKAPELVTDEM
VKQMRPGSVSLTLLLTKVALSKQLTVQRTMNPSMKNTVFSTMPLPISLVRLLALQPSPPMSLFLISKLWLAKDSHKQSLKMKP
CVKVLLIKVTLTYQLLKDLIVTTLISMI

SeqID 189

MKINKKYLAGSVAVLALSVCSYELGRHQAGQVKKESNRVSYIDGDQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSH
GDHYHYYNGKVPYDAIISEELLMKDPNYQLKDSDIVNEIKGGYVIKVDGKYYVYLKDAAHADNIRTKEEIKRQKQEHSHNHGG
GSNDQAVVAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAYWNGKQGSRPSSSSSYNANP
AQPRLSENHNLTVTPTYHQNQGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEK
RIARIIPLRYRSNHWVPDSRPEQPSPQSTPEPSPSPQPAPNPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLS
AETAAGIDSKLAKQESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVPSDKVKLVDD
ILAFLAPIRHPERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTHSHWIKKDSLSEAERAAAQ
AYAKEKGLTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKPEWFDEGLYEAP
KGYTLEDLLATVKYYVEHPNERPHSDNGFGNASDHVRKNKVDQDSKPDEDKEHDEVSEPTHPESDEKENHAGLNPSADNLYKP
STDTEETEEEAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDPSIRQNAMETLTGLKSSLLLGTKDNNTISAEVDSLLALL
KESQPAPIQ

SeqID 190
MKFSKKYIAAGSAVIVSLSLCAYALNQHRSQENKDNNRVSYVDGSQSSQKSENLTPDQVSQKEGIQAEQIVIKITDQGYVTSH
GDHYHYYNGKVPYDALFSEELLMKDPNYQLKDADIVNEVKGGYIIKVDGKYYVYLKDAAHADNVRTKDEINRQKQEHVKDNEK
VNSNVAVARSQGRYTTNDGYVFNPADIIEDTGNAYIVPHGGHYHYIPKSDLSASELAAAKAHLAGKNMQPSQLSYSSTASDNN
TQSVAKGSTSKPANKSENLQSLLKELYDSPSAQRYSESDGLVFDPAKIISRTPNGVAIPHGDHYHFIPYSKLSALEEKIARMV
PISGTGSTVSTNAKPNEVVSSLGSLSSNPSSLTTSKELSSASDGYIFNPKDIVEETATAYIVRHGDHFHYIPKSNQIGQPTLF
NNSLATPSPSLPINPGTSHEKHEEDGYGFDANRIIAEDESGFVMSHGDHNHYFFKKDLTEEQIKAAQKHLEEVKTSHNGLDSI
SSHEQDYPSNAKEMKDLDKKIEEKIAGIMKQYGVKRESIVVNKEKNAIIYPHGDHHHADPIDEHKPVGIGHSHSNYELFKPEE
GVAKKEGNKVYTGEELTNVVNLLKNSTFNNQNFTLANGQKRVSFSFPPELEKKLGINMLVKLITPDGKVLEKVSGKVFGEGVG
NIANFELDQPYLPGQTFKYTIASKDYPEVSYDGTFTVPTSLAYKMASQTIFYPFHAGDTYLRVNPQFAVPKGTDALVRVFDEF
HGNAYLENNYKVGEIKLPIPKLNQGTTRTAGNKIPVTFMANAYLDNQSTYIVEVPILEKENQTDKPSILPQFKRNKAQENLKI
DEKVEEPKTSEKVEKEKLSETGNSTSNSTLEEVPTVDPVQEKVAKPAESYGMKLENVLFNMDGTIELYLPSGEVIKKNMADFI
GEAPQGNGENKPSENGKVSTGTVENQPTENKPADSLPEAPNEKPVKPENSTDNGMLNPEGNVGSDPMLDPALEEAPAVDPVQF
KLEKFTASYGLGLDSVIFNMDGTIELRLPSGEVIKKNLSDLIA

SeqID 191
MKILFVAAEGAPFSKTGGLGDVIGALPKSLVKAGHEVAVILPYYDMVEAKFGNQIEDVLHFEVSVGWRRQYCGIKKTVLNGVI
FYFIDNQYYFFRGHVYGDFDDGERFAFFQLAAIEAMERIDFIPDLLHVHDYHTAMIPFLLKEKYRWIQAYEDIETVLTIHNLE
FQGQFSEGMLGDLFGVGFERYADGTLRWNNCLNWMKAGILYANRVSTVSPSYAHEIMTSQFGCNLDQILKMESGKVSGIVNGI
DADLYNPQTDALLDYHFNQEDLSGKAKNKAKLQERVGLPVRADVPLVGIVSRLTRQKGFDVVVESLHHILQEDVQIVLLGTGI
PAFEGAFSWFAQIYPDKLSTNITFDVKLAQEIYAACDLFLMPSRFEPCGLSQMMAMRYGTLPLVHEVGGLRDTVRAFNPIEGS
GTGFSFDNLSPYWLNWTFQTALDLYRNHPDIWRNLQKQAMESDFSWDTACKSYLDLYHSLVN

SeqID 192
MEKYFGEKQERFSPRKLSVGLVSATISSLFFMSVLASSSVDAQETAGVHYKYVADSELSSEEKKQLVYDIPTYVENDDETYYI
VYKLNSQNQLAELPNTGSKNERQALVAGASLAAMGILIFAVSKKKVKNKTVLHLVLVAGIGNGVLVSVHALENHLLLLNYNTDY
ELTSGEKLPLPKEISGYTYIGYIKEGKTTSESEVSNQKSSVATPTKQQKVDYNVTPNFVDHPSTVQAIQEQTPVSSTKPTEVQ
VVEKPFSTELINPRKEEKQSSDSQEQLAEHKNLETKKEEKISPKEKTGVNTLNPQDEVLSGQLNKPELLYREETMETKIDFQE
EIQENPDLAEGTVRVKQEGKLGKKVEIVRIFSVNKEEVSREIVSTSTTAPSPRIVEKGTKKTQVIKEQPETGVEHKDVQSGAI
VEPAIQPELPEAVVSDKGEPEVQPTLPEAVVTDKGETEVQPESPDTVVSDKGEPEQVAPLPEYKGNIEQVKPETPVEKTKEQG
PEKTEEVPVKPTEETPVNPNEGTTEGTSIQEAENPVQPAEESTTNSEKVSPDTSSKNTGEVSSNPSDSTTSVGESNKPEHNDS
KNENSEKTVEEVPVNPNEGTVEGTSNQETEKPVQPAEETQTNSGKIANENTGEVSNKPSDSKPPVEESNQPEKNGTATKPENS
GNTTSENGQTEPEPSNGNSTEDVSTESNTSNSNGNEEIKQENELDPDKKVEEPEKTLELRNVSDLELYSLSNGTYKQHISLEQ
VPSNPNSYFVKVKSSSFKDVYLPVASISEERKNDKILYKITAKVEKLQQEIESRYKDNFTFYLAKKGTEETTNFTSFSNLVKA
INQNPSGTYHLAASLNANEVELGPDERSYIKDTFTGRLIGEKDGKNYAIYNLKKPLFENLSGATVEKLSLKNVAISGKDDIGS
LANEAQNNTKIKQVHVDGVLAGERGIGGLLAKAEQSSITESSFKGRIINTYETTAAYNIGGMVGHLTGDKALLTKSKATVAIS
SNTNTSDQTVGGLAGLVDRDAQIQDSYAEGDINNVKHFGRVAGVAGNLWDRTSGDVRHAGSLTNVLSDVNVTNGNAITGYHYN
EMKVKDTFSSKANRVYNVTLVKDEVVSKESFEERGTMLDASQIASKKAEINPLILPTVEPLSTSGKKDSDFSKVAYYQAKRNI
TYKNIEKLLPFYNKATIVKYGNLVNENSLLYQKELLSAVMMKDNQVITDIVSNKQTANKLLLHYKDDLSEKLDLKYQNDFAKI
AEYSLGNTGLLYTPNQFLYDQTSIIKQVLPDLQKVDYHSEAIRKTLGISPNVKQTELYLEDQFAKTKQQLEDSLKKLLSADAG
LASANPVTEGYLVDKIKRNKEALLLGLTYLERWYNFSYGQVNVKDLVLYHLDFFGKGNASPLDTLIELGKSGFNNLLAKNNVL
TYGISLASQHGTTDLFSTLEHYRKVFLPNTSNNDWFKSETKAYIVEEKSTIEEVKTKQGLAGTKYSIGVYDRITSATWKYRNW
VLPLLTLPERSVFVISTMSSLGFGAYDRYRSSDHKAGKALNDFVEENARETAKRQRDHYDYWYRILDDNAREKLYRNILLYDA
YKFGDDNTVGKATEVADFDNPNPAMQHFFGPVGNKVGHNQHGAYATGDAVYYMGYRMLDKDGAITYTHEMTHDSDQDIYLGGY
GRRSGLGPEFFAKGLLQAPDHPDDATITINSILKHSKSDSTESRRLQVLDPTTRFNNADDLKQYVHNMFDVVYMLEYLEGNSI
LKLDTNQKQQLLRKVTNEYHPDPDGNKVYATNVVRNLTVEEVERLRSFNDLIDNNILSSREYASGKYERNGYFTIKLFAPIYA
ALSNDIGTPGDLMGRRIAYELLAAKGFKDGMVPYISNQYEEEAKQKGKTINLYGKTRGLVTDDLVLEKVFNNQYHTWSEFKKA
MYQERQDQFDRLNKVTFNDTTQPWQTFAKKTTSSVDELQKLMDVAVRKDAEHNYYHWNNYNPDIDSEVHKLKRAIFKAYLDQI
NDFRSSIFENKK

SeqID 193
MKINKKYLAGSVAVLALSVCSYELGRYQAGQDKKESNRVAYIDGDQAGQKAENLTPDEVSKREGINAEQIVIKITDQGYVTSH
GDHYHYYNGKVPYDAIISEELLMKDPNYQLKDSDIVNEIKGGYVIKVNGKYYVYLKDAAHADNIRTKEEIKRQKQERSHNHNS
RADNAVAAARAQGRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAYWNGKQGSRPSSSSSYNANPA
QPRLSENHNLTVTPTYHQNQGENISSLLRELYAKPLSERHVESDGLIFDPAQITSRTARGVAVPHGNHYHFIPYEQMSELEKR
IARIIPLRYRSNHWVPDSRPEEPSPQPTPEPSPSPQPAPSNPIDEKLVKEAVRKVGDGYVFEENGVSRYIPAKDLSAETAAGI

DSKLAKQESLSHKLGTKKTDLPSSDREFYNKAYDLLARIHQDLLDNKGRQVDFEALDNLLERLKDVSSDKVKLVEDILAFLAF
IRHPERLGKPNAQITYTDDEIQVAKLAGKYTTEDGYIFDPRDITSDEGDAYVTPHMTHSHWIKKDSLSEAERAAAQAYAKEKG
LTPPSTDHQDSGNTEAKGAEAIYNRVKAAKKVPLDRMPYNLQYTVEVKNGSLIIPHYDHYHNIKFEWFDEGLYEAPKGYTLEC
LLATVKYYVEHPNERPHSDNGFGNASDHVQRNKNGQADTNQTEKPSEEKPQTEKPEEETPREEKPQSEKPESPKPTEEPEESF
EESEEPQVETEKVEEKLREAEDLLGKIQDPIIKSNAKETLTGLKNNLLFGTQDNNTIMAEAEKLLALLKESK

SeqID 194
LILSVCSYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSHGDHYHYYNGKVPYDA
IISEELLMKDPNYKLKDEDIVNEVKGGYVIKVDGKYYVYLKDAAHADNVRTKEEINRQKQEHSQHREGGTPRNDGAVALARSC
GRYTTDDGYIFNASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPSVSN
PGTTNTNTSNNSNTNSQASQSNDIDSLLKQLYKLPLSQRHVESDGLVFDPAQITSRTARGVAVPHGDHYHFIPYSQMSELEEF
IARIIPLRYRSNHWVPDSRPEQPSPQPTPEPSPGPQPAPNLKIDSNSSLVSQLVRKVGEGYVFEEKGISRYVFAKDLPSETVK
NLESKLSKQESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLERLNDESTNKEKLVDDLLAFL
APITHPERLGKPNSQIEYTEDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGHSHWIGKDSLSDKEKVAAQAYTKE
KGILPPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFAWFDDHTYKAPNGYTL
EDLFATIKYYVEHPDERPHSNDGWGNASEHVLGKKDHSEDPNKNFKADEEPVEETPAEPEVPQVETEKVEAQLKEAEVLLAKV
TDSSLKANATETLAGLRNNLTLQIMDNNSIMAEAEKLLALLKGSNPSSVSKEKIN

SeqID 195
MPVEIKTTKEIHPKIYAYTTPTVTSNEGWIKIGYTERDVTQRIKEQTHTAHIATDVLWTGDAAYTEEPDKGKTFKDHDFHHFL
SFHDVERRPKTEWFYFNGTPEKSKNLFDKFVQHDLSGYQPGKGQDYTLRQEQEEAVAKTLAYFQEHAGGKFLWNAKPRFGKTL
STYDLARRMEAVNVLIVTNRPAIANSWYDDFETFIAGQTTYKFVSESDSLKSRPILSRQEFLGILADDVRQLAFISLQDLKGS
VYLGGEHDKLKWVTDLHWDLLVIDEAHEGVDTPKTDQAFNKIRRNFTLHLSGTSPKALAKGDFTEEQIYNWSYADEQAAKYSW
SLEQEEENPYESLPQLNLFTYQMSQMIGEKLEKGAQIDGENIDYVFDLSEFFATDDKGKFIHEHDVRNWLDTLSSNEKYPFST
KELRNELKHTFWLLERVASAKALKALLEEHPIYENYEIVLAAGDGRMSEEDDKVKLKSLDLVRKAIAENDKTITLSVGQLTTG
VTIPEWTGVLMLSNLKSPALYMQAAFRAQNPYSWSDNKGNHFRKERAYVFDFAPERTLILFDEFANNLLLVTAAGRGTSATRE
ENIRELLNFFPIIAEDRAGKMVEIDAKAVLTTPRQIKAREVLKRGFMSNLLFDNISGIFQASQTVLDILNELPVEKEGKVQDS
SDLLDFSDVTVDDEGNAVVDHEIVVNQQMRLFGEKVYGLGESVAELVTKDEERTQKQLVNDLSKTVSSVIVEELKADYSLKTR
ETEQIKKQITATLENEIRKNDIERKISEAHIKQELQQQLKEANDKAQKDKIQEDLEKRLEENKLIHKEKLEQTLKKEVEKMPE
KFIEQVEIKRVEQLKQSAQDEIRDHLRGFARTIPSFIMAYGDQTLTLDNFDAFVPEHVFYEVTGITIDQFRYLRDGGQDFAGH
LFDKATFDEAIQEFLRKKKELADYFKDQKEDIFDYIPPQKTNQIFTPKRVVKRMVDDLEKENPGIFDDPSKTFIDLYMKSGLY
IAELVKRLYNSNGLKEAFPNPEERLKHILEKQVYGFAPSEIIYNISTNFIFGNLSKDISRKNFVLADTIPAAKEGSIQKLVDS
YFENN

SeqID 196
MKKILIVDDEKPISDIIKFNMTKEGYEVVTAFNGREALEQFEAEQPDIIILDLMLPEIDGLEVAKTIRKTSSVPILMLSAKDS
EFDKVIGLELGADDYVTKPFSNRELQARVKALLRRSQPMPVDGQEADSKPQPIQIGDLEIVPDAYVAKKYGEELDLTHREFEL
LYHLASHTGQVITREHLLETVWGYDYFGDVRTVDVTVRRLREKIEDTPSRPEYILTRRGVGYYMRNNA

SeqID 197
MKKKFLAFLLILFPIFSLGIAKAETIKIVSDTAYAPFEFKDSDQTYKGIDVDIINKVAEIKGWNIQMSYPGFDAAVNAVQAGQ
ADAIMAGMTKTKEREKVFTMSDTYYDTKVVIATTKSHKISKYDQLTGKTVGVKNGTAAQRFLETIKDKYGFTIKTFDTGDLMN
NSLSAGAIDAMMDDKPVIEYAINQGQDLHIEMDGEAVGSFAFGVKKGSKYEHLVTEFNQALSEMKKDGSLDKIIKKWTASSSS
AVPTTTTLAGLKAIPVKAKYIIASDSSFAPFVFQNSSNQYTGIDMELIKAIAKDQGFEIEITNPGFDAAISAVQAGQADGIIA
GMSVTDARKATFDFSESYYTANTILGVKESSNIASYEDLKGKTVGVKNGTASQTFLTENQSKYGYKIKTFADGSSMYDSLNTG
AIDAVMDDEPVLKYSISQGQKLKTPISGTPIGETAFAVKKGANPELIEMFNNGLANLKANGEFQKILDKYLASESSTASTSTV
DETTLWGLLQNNYKQLLSGLGITLALALISFAIAIVIGIIFGMFSVSPYKSLRVISEIFVDVIRGIPLMILAAFIFWGIPNFI
ESITGQQSPINDFVAGTIALSLNAAAYIAEIVRGGIQAVPVGQMEASRSLGISYGKTMRKIILPQATKLMLPNFVNQFVIALK
DTTIVSAIGLVELFQTGKIIIARNYQSFKMYAILAIFYLVIITLLTRLAKRLEKRIR

SeqID 198
MAFESLTERLQNVFKNLRKKGKISESDVQEATKEIRLALLEADVALPVVKDFIKKVRERAVGHEVIDTLNPAQQIIKIVDEEL
TAVLGSDTAEIIKSPKIPTIIMMVGLQGAGKTTFAGKLANKLKKEENARPLMIAADIYRPAAIDQLKTLGQQIDVPVFALGTE
VPAVEIVRQGLEQAQTNHNDYVLIDTAGRLQIDELLMNELRDVKALAQPNEILLVVDAMIGQEAANVAREFNAQLEVTGVILT
KIDGDTRGGAALSVRHITGKPIKFTGTGEKITDIETFHPDRMSSRILGMGDMLTLIEKASQEYDEQKALEMAEKMRENTFDFN
DFIDQLDQVQNMGPMEDLLKMIPGMANNPALQNMKVDERQIARKRAIVSSMTPEERENPDLLNPSRRRIAAGSGNTFVEVNK
FIKDFNQAKQLMQGVMSGDMNKMMKQMGINPNNLPKNMPNMGGMDMSALEGMMGQGGMPDLSALGGAGMPDMSQMFGGGLKGK
IGEFAMKQSMKRMANKMKKAKKKRK

SeqID 199
MSQIWTKEKFISQVQGGVIVSCQALPGEALYNEEFSLMPFMAKAALEAGAVGIRANSVRDIKAIQKVVDLPIIGIIKRDYPPQ
EPYITATMKEVDELVECGTTVIAFDATLRPRYDGLVVSEFIKKIKEKYPNQLLMADVSNLDEGLYAFKSGVDFVGTTLSGYTS
TSVQSDEPDFELMKKLADFNIPVIAEGKIHYPEQLKKAYSLGVTSVVIGGAITRPKEIAQRFINVIK

SeqID 200
MRYLTAGESHGPRLTAIIEGIPAGLPLTAEDINEDLRRRQGGYGRGGRMKIENDQVVFTSGVRHGKTTGAPITMDVINKDHQK
WLDIMSAEDIEDRLKSKRKITHPRPGHADLVGGIKYRFDDLRNSLERSSARETTMRVAVGAVAKRLLAELDMEIANHVVVFGG
KEIDVPENLTVAEIKQRAAQSEVSIVNQEREQEIKDYIDQIKRDGDTIGGVVETVVGGVPVGLGSYVQWDRKLDARLAQAVVS

INAFKGVEFGLGFEAGYRKGSQVMDEILWSKEDGYTRRTNNLGGFEGGMTNGQPIVVRGVMKPIPTLYKPLMSVDIETHEPYK
ATVERSDPTALPAAGMVMEAVVATVLAQEILEKFSSDNLEELKEAVAKHRDYTKNY

SeqID 201
MVVMNRIRVSKRVEKKLAKGLVLLEASDLENVNLKDQEVEVQGQEGNFLGTAYLSQQNKGLGWFISKDKVAFNQAFFETLFRK
AKEKRNAYYQDDLTTAFRLFNQEGDGFGGLTVDLYGDYAVFSWYNSYVYQIRQTISEAFRQVFPEVLGAYEKIRFKGLDYESA
HVYGQEAPDFFNVLENGVLYQVFMNDGLMTGIFLDQHEVRGSLVDGLAMGKSLLNMFSYTAAFSVAAAMGGASHTTSVDLAKR
SRELSQAHFQANGLSTDEHRFIVMDVFEYFKYAKRKDLTYDVIVLDPPSFARNKKQTFSVAKDYHKLISQSLEILNPGGIIIA
STNAANVSRQKFTEQIDKGFAGRSYQILNKYGLPADFAYNKKDESSNYLKVISMKVSK

SeqID 202
MTKTLKRPEVLSPAGTLEKLKVAVQYGADAVFIGGQAYGLRSRAGNFTFEQMEEGVQFAAKYGAKVYVAANMVMHEGNEAGAG
EWFRKLRDIGIAAVIVSDPALIMIAVTEAPGLEIHLSTQASATNYETLEFWKELGLTRVVLAREVSMEELAEIRKRTDVEIEA
FVHGAMCISYSGRCTLSNHMSRDANRGGCSQSCRWKYDLYDMPFGKERKSLQGEIPEEFSMSAVDMSMIDHIPDMIENGVDS
LKIEGRMKSIHYVSTVTNCYKAAVDAYLESPEKFEAIKQDLVDEMWKVAQRELATGFYYGTPSENEQLFGARRKIPEYKFVAE
VVSYDDAAQTATIRQRNVINEGDQVEFYGPGFRHFETYIEDLHDAKGNKIDRAPNPMELLTIKVPQPVQSGDMVRALKEGLIN
LYKEDGTSVTVRA

SeqID 203
MNTYQLNNGVEIPVLGFGTFKAKDGEEAYRAVLEALKAGYRHIDTAAIYQNEESVGQAIKDSGVPREEMFVTTKLWNSQQTYE
QTRQALEKSIEKLGLDYLDLYLIHWPNPKPLRENDAWKTRNAEVWRAMEDLYQEGKIRAIGVSNFLPHHLDALLETATIVPAV
NQVRLAPGVYQDQVVAYCREKGILLEAWGPFGQGELFDSKQVQEIAANHGKSVAQIALAWSLAEGFLPLPKSVTTSRIQANLD
CFGIELSHEERETLKTIAVQSGAPRVDDVDF

SeqID 204
LSEKSREEEKLSFKEQILRDLEKVKGYDEVLKEDEAVVRTPANEPSTEELMADSLSTVEEIMRKAPTVPTHPSQGVPASPADE
IQRETPGVPSHPSQDVPSSPAEESGSRPGPGPVRPKKLEREYNETPTRVAVSYTTAEKKAEQAGPETPTPATETVDIIRDTSR
RSRREGAKPVKPKKEKKSHVKAFVISFLVFLALLSAGGYFGYQYVLDSLLPIDANSKKYVTVGIPEGSNVQEIGTTLEKAGLV
KHGLIFSFYAKYKNYTDLKAGYYNLQKSMSTEDLLKELQKGGTDEPQEPVLATLTIPEGYTLDQIAQAVGQLQGDFKESLTAE
AFLAKVQDETFISQAVAKYPTLLESLPVKDSGARYRLEGYLFPATYSIKESTTIESLIDEMLAAMDKNLSPYYSTIKSKNLTV
NELLTIASLVEKEGAKTEDRKLIAGVFYNRLNRDMPLQSNIAILYAQGKLGQNISLAEDVAIDTNIDSPYNVYKNVGLMPGPV
DSPSLDAIESSINQTKSDNLYFVADVTEGKVYYANNQEDHDRNVAEHVNSKLN

SeqID 205
MKQERFPLVSDDEVMLTEMPVMNLYDESDLISNIKGEYRDKNYLEWAPIAEEKPVKPIEKQVEKPKKAPLGVKKEGKSYAEVA
REEARADLKKKRSANYLTQDFSLARRHSQPSLVRQGNQPTAPFQKENPGEFVKYSQKLTQSHYILAEEVHSIPTKNEEVSAPA
PKKNNYDFLKKSQIYNKKSKQTEQERRVAQELNLTRMTE

SeqID 206
MKKSKSKYLTLAGLVLGTGVLLSACGNSSTASKTYNYVYSSDPSSLNYLAENRAATSDIVANLVDGLLENDQYGNIIPSLAED
WTVSQDGLTYTYKLRKDAKWFTSEGEEYAPVTAQDFVTGLQYAADKKSEALYLVQDSVAGLDDYITGKTSDFSTVGVKALDDQ
TVQYTLVKPELYWNSKTLATILFPVNADFLKSKGDDFGKADPSSILYNGPFLMKALVSKSAIEYKKNPNYWDAKNVFVDDVKL
TYYDGSDQESLERNFTAGAYTTARLFPNSSSYEGIKEKYKNNIIYSMQNSTSYFFNFNLDRKSYNYTSKTSDIEKKSTQEAVL
NKNFRQAINFAFDRTSYGAQSEGKEGATKILRNLVVPPNFVSIKGKDFGEVVASKMVNYGKEWQGINFADGQDPYYNPEKAKA
KFAEAKKELEAKGVQFPIHLDKTVEVTDKVGIQGVSSIKQSIESVLGSDNVVIDIQQLTSDEFDSSGYFAQTAAQKDYDLYHG
GWGPDYQDPSTYLDIFNTNSGGFLQNLGLEPGEANDKAKAVGLDVYTQMLEEANKEQDPAKRYEKYADIQAWLIDSSLVLPSV
SRGGTPSLRRTVPFAAAYGLTGTKGVESYKYLKVQDKIVTTDEYAKAREKWLKEKEESNKKAQEELAKHVK

SeqID 207
VEQHSDVCYIFYRRERLKTKIGLASICLLGLATSHVAANETEVAKTSQDTTTASSSSEQNQSSNKTQTSAEVQTNAAAHWDGD
YYVKDDGSKAQSEWIFDNYYKAWFYINSDGRYSQNEWHGNYYLKSGGYMAQNEWIYDSNYKSWFYLKSDGAYAHQEWQLIGNK
WYYFKKWGYMAKSQWQGSYFLNGQGAMMQNEWLYDPAYSAYFYLKSDGTYANQEWQKVGGKWYYFKKWGYMARNEWQGNYYLT
GSGAMATDEVIMDGTRYIFAASGELKEKKDLNVGWVHRDGKRYFFNNREEQVGTEHAKKVIDISEHNGRINDWKKVIDENEVD
GVIVRLGYSGKEDKELAHNIKELNRLGIPYGVYLYTYAENETDAESDAKQTIELIKKYNMNLSYPIYYDVENWEYVNKSKRAP
SDTGTWVKIINKYMDTMKQAGYQNVYVYSYRSLLQTRLKHPDILKHVNWVAAYTNALEWENPHYSGKKGWQYTSSEYMKGIQG
RVDVSVWY

SeqID 208
MAKEPWQEDIYDQEESRAERRHRNHGGADRMANRILTILASIFFVIVVVMVIVLIYLSSGGSNRTAALKGFHDSDASVVQISS
SSSSQPEQSSEPESTSSSSEEAANPEGTIKVLAGEGEAAIAARAGISIAQLEALNPGHMATGSWFANPGDVIKIK

SeqID 209
MPITSLEIKDKTFGTRFRGFDPEEVDEFLDIVVRDYEDLVRANHDKNLRIKSLEERLSYFDEIKDSLSQSVLIAQDTAERVKQ
AAHERSNNIIHQAEQDAQRLLEEAKYKANEILRQATDNAKKVAVETEELKNKSRVFHQRLKSTIESQLAIVESSDWEDILRPT
ATYLQTSDEAFKEVVSEVLGEPIPAPIEEEPIDMTRQFSQAEMAELQARIEVADKELSEFEAQIKQEVEAPTPVVSPQVEEEP
LLIQLAQCMKNQK

SeqID 210

MSLKDRFDRFIDYFTEDEDSSLPYEKRDEPVFTSVNSSQEPALPMNQPSQSAGTKENNITRLHARQQELANQSQRATDKVIID
VRYPRKYEDATEIVDLLAGNESILIDFQYMTEVQARRCLDYLDGACHVLAGNLKKVASTMYLLTPVNVIVNVEDIRLPDEDQQ
GEFGFDMKRNRVR

SeqID 211
MSDLKKYEGVIPAFYACYDDQGEVSPERTRALVQYFIDKGVQGLYVNGSSGECIYQSVEDRKLILEEVMAVAKGKLTIIAHVA
CNNTKDSMELARHAESLGVDAIATIPPIYFRLPEYSVAKYWNDISSAAPNTDYVIYNIPQLAGVALTPSLYTEMLKNPRVIGV
KNSSMPVQDIQTFVSLGGEDHIVFNGPDEQFLGGRLMGARAGIGGTYGAMPELFLKLNQLIADKDLETARELQYAINAIIGKL
TSAHGNMYGVIKEVLKINEGLNIGSVRSPLTPVTEEDRPVVEAAAALIRETKERFL

SeqID 212
MNKRGLYSKLGISVVGISLLMGVPTLIHANELNYGQLSISPIFQGGSYQLNNKSIDISSLLLDKLSGESQTVVMKFKADKPNS
LQALFGLSNSKAGFKNNYFSIFMRDSGEIGVEIRDAQKGINYLFSRPASLWGKHKGQAVENTLVFVSDSKDKTYTMYVNGIEV
FSETVDTFLPISNINGIDKATLGAVNREGKEHYLAKGSIDEISLFNKAISDQEVSTIPLSNPFQLIFQSGDSTQANYFRIPTL
YTLSSGRVLSSIDARYGGTHDSKSKINIATSYSDDNGKTWSEPIFAMKFNDYEEQLVYWPRDNKLKNSQISGSASFIDSSIVE
DKKSGKTILLADVMPAGIGNNNANKADSGFKEINGHYYLKLKKNGDNDFRYTVRENGVVYNETTNKPTNYTINDKYEVLEGGK
SLTVEQYSVDFDSGSLRERHNGKQVPMNVFYKDSLFKVTPTNYIAMTTSQNRGESWEQFKLLPPFLGEKHNGTYLCPGQGLAL
KSSNRLIFATYTSGELTYLISDDSGQTWKKSSASIPFKNATAEAQMVELRDGVIRTFFRTTTGKIAYMTSRDSGETWSKVSYI
DGIQQTSYGTQVSAIKYSQLIDGKEAVILSTPNSRSGRKGGQLVVGLVNKEDDSIDWKYHYDIDLPSYGYAYSAITELPNHHI
GVLFEKYDSWSRNELHLSNVVQYIDLEINDLTK

SeqID 213
MNRSVQERKCRYSIRKLSVGAVSMIVGAVVFGTSPVLAQEGASEQPLANETQLSGESSTLTDTEKSQPSSETELSGNKQEQER
KDKQEEKIPRDYYARDLENVETVIEKEDVETNASNGQRVDLSSELDKLKKLENATVHMEFKPDAKAPAFYNLFSVSSATKKDE
YFTMAVYNNTATLEGRGSDGKQFYNNYNDAPLKVKPGQWNSVTFTVEKPTAELPKGRVRLYVNGVLSRTSLRSGNFIKDMPDV
THVQIGATKRANNTVWGSNLQIRNLTVYNRALTPEEVQKRSQLFKRSDLEKKLPEGAALTEKTDIFESGRNGNPNKDGIKSYR
IPALLKTDKGTLIAGADERRLHSSDWGDIGMVIRRSEDNGKTWGDRVTITNLRDNPKASDPSIGSPVNIDMVLVQDPETKRIF
SIYDMFPEGKGIFGMSSQKEEAYKKIDGKTYQILYREGEKGAYTIRENGTVYTPDGKATDYRVVVDPVKPAYSDKGDLYKGDQ
LLGNIYFTTNKTSPFRIAKDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYTTNNVS
HLDGSQSSRVIYSDDHGKTWHAGEAVNDNRQVDGQKIHSSTMNNRRAQNTESTVVQLNNGDVKLFMRGLTGDLQVATSKDGGV
TWEKDIKRYPQVKDVYVQMSAIHTMHEGKEYIILSNAGGPKRENGMVHLARVEENGELTWLKHNPIQKGEFAYNSLQELGNGE
YGILYEHTEKGQNAYTLSFRKFNWEFLSKNLISPTEANRDGQRRDGQRSYWLGVRLRSIGQQGSNPSIGKWNSDFPNPVQDLV
VCSRGRYRTGNYWYSRKHRKYASSCKSSRCQSSWRSKWQSSGASSRIYRGSWYRASCSNRRVGIFACNSYYKKRLYLQSSSCS
AGTSNRKQGEPPSFTRTNSFLPWSVYAREKERT

SeqID 214
MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEED
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAVRIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSL
TQTNSMLGSVHYLSPEQARGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQKPLPSVIAENPSVPQALENVIIK
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTEHKSIKNPSQAVTEETYQ
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDP
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATQ
IVLTVAKKATTIQLGNYIGRNSTEVISELKQKKVPENLIKIEEEESSESEPGTIMKQSPGAGTTYDVSKPTQIVLTVAKKVTS
VAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSAEGMVVEQSPRAGEKVDLNKTRVKISIYKPKTTSATP

SeqID 215
MTKLIFMGTPDFSATVLKGLLTDDRYEILAVVTQPDRAVGRKKVIQETPVKQAAKEAGLSIYQPEKLSGSPEMEDLMKLGADG
IVTAAFGQFLPSKLLDSMDFAVNVHASLLPRHRGGAPIHYALIQGDEEAGVTIMEMVKEMDAGDMISRRSIPITDEDNVGTLF
EKLALVGRDLLLDTLPAYIAGDIKPEPQDTSQVTFSPNIKPEEEKLDWNKTNRQLFNQIRGMNPWPVAHTFLKGDRFKIYEAL
PVEGQGNPGEILSIGKKELIVATAEGALSLKQVQPAGKPKMDIASFLNGVGRTLTVGERFGD

SeqID 216
VFRRLGQDFQLRKVKKILKQINALKGKMSSLSDQELVAKTVEFRQRLSEGESLDDILVEAFAVVREADKRILGMFPYDVQVMG
AIVMHYGNVAEMNTGEGKTLTATMPVYLNAFSGEGVMVVTPNEYLSKRDAEEMGQVYRFLGLTIGVPFTEDPKKEMKAEEKKL
IYASDIIYTTNSNLGFDYLNDNLASNEEGKFLRPFNYVIIDEIDDILLDSAQTPLIIAGSPRVQSNYYAIIDTLVTTLVEGED
YIFKEEKEEVWLTTKGAKSAENFLGIDNLYKEEHASFARHLVYAIRAHKLFTKDKDYIIRGNEMVLVDKGTGRLMEMTKLQGG
LHQAIEAKEHVKLSPETRAMASITYQSLFKMFNKISGMTGTGKVAEKEFIETYNMSVVRIPTNRPQRIDYPDNLYITLPEKV
YASLEYIKQYHAKGNPLLVFVGSVEMSQLYSSLLFREGIAHNVLNANNAAREAQIISESGQMGAVTVATSMAGRGTDIKLGKG
VAELGGLIVIGTERMESQRIDLQIRGRSGRQGDPGMSKFFVSLEDDVIKKFGPSWVHKKYKDYQVQDMTQPEVLKGRKYRKLV
EKAQHASDSAGRSARRQTLEYAESMNIQRDIVYKERNRLIDGSRDLEDVVVDIIERYTEEVAADHYASRELLFHFIVTNISFH
VKEVPDYIDVTDKTAVRSFMKQVIDKELSEKKELLNQHDLYEQFLRLSLLKAIDDNWVEQVDYLQQLSMAIGGQSASQKNPIV
EYYQEAYAGFEAMKEQIHADMVRNLLMGLVEVTPKGEIVTHFP

SeqID 217
MTETVEDKVSHSITGLDILKGIVAAGAVISGTVATQTKVFTNESAVLEKTVEKTDALATNDTVVLGTISTSNSASSTSLSASE
SASTSASESASTSASTSASTSASESASTSASTSISASSTVVGSQTAAATEATAKKVEEDRKKPASDYVASVTNVNLQSYAKRR
KRSVDSIEQLLASIKNAAVFSGNTIVNGAPAINASLNIAKSETKVYTGEGVDSVYRVPIYYKLKVTNDGSKLTFTYTVTYVNP
KTNDLGNISSMRPGYSIYNSGTSTQTMLTLGSDLGKPSGVKNYITDKNGRQVLSYNTSTMTTQGSGYTWGNGAQMNGFFAKKG

```
YGLTSSWTVPITGTDTSFTFTPYAARTDRIGINYFNGGGKVVESSTTSQSLSQSKSLSVSASQSASASASASTSASASASTSASA
SASTSASASASTSASVSASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASAS
TSASESASTSASASASTSASASASTSASGSASTSTSASASTSASASASTSASASASISASESASTSASESASTSTSASASTSA
SESASTSASASASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASVSAS
TSASASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSA
SASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASAS
SASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSASESASTSASASASTSASASAST
ASTSASASASTSASASASTSASASASISASESASTSASASASTSASVSASTSASASASTSASESASTSASASASTSASESAST
SASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASESASTSTSASASTSASESASTSASASASTSAS
ASASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASAST
SASASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASVSASTSASESASTSASASASTSAS
ASASTSASESASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASA
STSASASASTSASASASTSASASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASVSASTS
ASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASA
SASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASESASTSASASAS
TSASVSASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSF
SASASTSASASASTSASASASTSASASASTSASASASTSASASASISASESASTSASESASTSTSASASTSASESASTSAS
ASASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASVSAST
SASASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSAS
ASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASF
STSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSASASASTSASASASTS
ASASASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASF
STSASASASTSASASASTSASASASISASESASTSASASASTSASVSASTSASASASTSASESASTSASASASTSASESAST
ASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASESASTSTSASASTSASESASTSASASASTSASI
SASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTS
ASASTSASASASTSASASASTSASASTSASESASTSASASASTSASESASTSASASASTSASASASTSASVSASTSASI
SASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASAS
TSASASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASASASISASESI
STSASESASTSTSASASTSASESASTSASASASTSASASASTSASASASTSASTSASESASTSASASASTSASASASTSAS
ASASTSASASASTSASASASTSASVSASTSASASASTSASASASTSASESASTSASASTSASESASTSASASASTSASASAS
SASASASTSASESASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSAS
ESASTSASASASTSASASASTSASGSASTSTSASASTSASASASTSASASASISASESASTSASESASTSTSASASTSASESI
STSASASASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASVSASTSA
ASASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASI
STSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASTSASASAST
ASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASI
SASTSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASAST
ASASASTSASASASTSASASASISASESASTSASASASTSASVSASTSASASASTSASESASTSASASASTSASESASTSASI
SASTSASASASISASESASTSASASASTSASASASTSASASASTSASESASTSTSASASTSASESASTSASASASTSASASAS
TSASASASTSASASASTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASI
SASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASVSASTSASESASTSASASASTSASASAS
TSASESASTSASASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSI
SASASTSASASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASVSASTSASAS
ASTSASASASISASESASTSASASASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASAS
SASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSASASASTSASASASISASESASTSAI
ASASTSASASASTSASASASTSASESASTSTSASASTSASESASTSASASASTSASASASTSASASASTSASASASTSASAS
SASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASESASTSAI
ASASTSASASASTSASASASTSASVSASTSASESASTSASASASTSASASASTSASESASTSASASASTSASESI
STSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASAST
ASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASVSASTSASASASTSASASASISASESASTSASI
SASTSASASASTSASASASTSASASASISASESASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASAI
TSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSASASASTSVSNSANHSNSQVGNTS(
STGKSQKELPNTGTESSIGSVLLGVLAAVTGIGLVAKRRKRDEEE
```

SeqID 218
```
MSNEKNTNTNVEKKDATVVAHEIKGELTYEDKVIQKIIGLSLENVSGLLGIDGGFFSNLKEKIVNSDDVTSGVNVEVGKTQVZ
VDLNVIVEYQKNVPALYSEIREIVSSEVAKMTDLEIVEINVNVVDIKTKEQHEADSVSLQDRVSDVAESTGEFTSEQFEKAKS
GLGSGFSTVQEKVSEGVEAVKGAANGVVSHENTRVN
```

SeqID 219
```
MTKEKNVILTARDIVVEFDVRDKVLTAIRGVSLELVEGEVLALVGESGSGKSVLTKTFTGMLEENGRIAQGSIDYRGQDLTAI
SSHKDWEQIRGAKIATIFQDPMTSLDPIKTIGSQITEVIVKHQGKTAKEAKELAIDYMNKVGIPDADRRFNEYPFQYSGGMRC
RIVIAIALACRPDVLICDEPTTALDVTIQAQIIDLLKSLQNEYHFTTIFITHDLGVVASIADKVAVMYAGEIVEYGTVEEVF?
DPRHPYTWSLLSSLPQLADDKGDLYSIPGTPPSLYTDLKGDAFALRSDYAMQIDFEQKAPQFSVSETHWAKTWLLHEDAPKVJ
KPAVIANLHDKIREKMGFAHLAD
```

SeqID 220

```
MKKNRVFATAGLVLLAAGVLAACSSSKSSDSSAPKAYGYVYTADPETLDYLISSKNSTTVVTSNGIDGLFTNDNYGNLAPAVA
EDWEVSKDGLTYTYKIRKGVKWFTSDGEEYAEVTAKDFVNGLKHAADKKSEAMYLAENSVKGLADYLSGTSTDFSTVGVKAVD
DYTLQYTLNQPEPFWNSKLTYSIFWPLNEEFETSKGSDFAKPTDPTSLLYNGPFLLKGLTAKSSVEFVKNEQYWDKENVHLDT
INLAYYDGSDQESLERNFTSGAYSYARLYPTSSNYSKVAEEYKDNIYYTQSGSGIAGLGVNIDRQSYNYTSKTTDSEKVATKK
ALLNKDFRQALNFALDRSAYSAQINGKDGAALAVRNLFVKPDFVSAGEKTFGDLVAAQLPAYGDEWKGVNLADGQDGLFNADK
AKAEFAKAKKALEADGVQFPIHLDVPVDQASKNYISRIQSFKQSVETVLGVENVVVDIQQMTSDEFLNITYYAANASSEDWDV
SGGVSWGPDYQDPSTYLDILKTTSSETTKTYLGFDNPNSPSVVQVGLKEYDKLVDEAARETSDLNVRYEKYAAAQAWLTDSSL
FIPAMASSGAAPVLSRIVPFTGASAQTGSKGSDVYFKYLKSQDKVVTKEEYEKAREKWLKEKAESNEKAQKELASHVK


SeqID 221
MEINVSKLRTDLPQVGVQPYRQVHAHSTGNPHSTVQNEADYHWRKDPELGFFSHIVGNGCIMQVGPVDNGAWDVGGGWNAETY
AAVELIESHSTKEEFMTDYRLYIELLRNLADEAGLPKTLDTGSLAGIKTHEYCTNNQPNNHSDHVDPYPYLAKWGISREQFKH
DIENGLTIETGWQKNDTGYWYVHSDGSYPKDKFEKINGTWYYFDSSGYMLADRWRKHTDGNWYWFDNSGEMATGWKKIADKWY
YFNEEGAMKTGWVKYKDTWYYLDAKEGAMVSNAFIQSADGTGWYYLKPDGTLADKPEFTVEPDGLITVK


SeqID 222
MKKKYWTLAILFFCLFNNSVTAQEIPKNLDGNITHTQTSESFSESDEKQVDYSNKNQEEVDQNKFRIQIDKTELFVTTDKHLE
KNCCKLELEPQINNDIVNSESNNLLGEDNLDNKIKENVSHLDNRGGNIEHDKDNLESSIVRKYEWDIDKVTGGGESYKLYSKS
NSKVSIAILDSGVDLQNTGLLKNLSNHSKNYVPNKGYLGKEEGEEGIISDIQDRLGHGTAVVAQIVGDDNINGVNPHVNINVY
RIFGKSSASPDWIVKAIFDAVDDGNDIINLSTGQYLMIDGEYEDGTNDFETFLKYKKAIDYANQKGVIIVAALGNDSLNVSNQ
SDLLKLISSRKKVRKPGLVVDVPSYFSSTISVGGIDRLGNLSDFSNKGDSDAIYAPAGSTLSLSELGLNNFINAEKYKEDWIF
SATLGGYTYLYGNSFAAPKVSGAIAMIIDKYKLKDQPYNYMFVKKFWKKHYQ


SeqID 223
MKKDELFEGFYLIKSADLRQTRAGKNYLAFTFQDDSGEIDGKLWDAQPHNIEAFTAGKVVHMKGRREVYNNTPQVNQITLRLP
QAGEPNDPADFKVKSPVDVKEIRDYMSQMIFKIENPVWQRIVRNLYTKYDKEFYSYPAAKTNHHAFETGLAYHTATMVRLADA
ISEVYPQLNKSLLYAGIMLHDLAKVIELTGPDQTEYTVRGNLLGHIALIDSEITKTVMELGIDDTKEEVVLLRHVILSHHGLL
EYGSPVRPRIMEAEIIHMIDNLDASMMMMSTALALVDKGEMTNKIFAMDNRSFYKPDLD


SeqID 224
VTILGKDTVQQSAKGESVTQEATPEYKLENTPGGDKGGNTGSSDANANEGGGSQAGGSAHTGSQNSAQSQASKQLATEKESAK
NAIEKAAKDKQDEIKGAPLSDKEKAELLARVEAEKQAALKEIENAKTMEDVKEAETIGVQAIAMVTVPKRPVAPNAAPKTTSA
PQATAGTMQDVTYQSPAGKQLPNTGSASSAALASLGLVVATSGFALLGRKTRRRK


SeqID 225
MNADDTVTIYDVAREAGVSMATVSRVVNGNKNVKENTRKKVLEVIDRLDYRPNAVARGLASKKTTTVGVVIPNITNGYFSSLA
KGIDDIAEMYKYNIVLANSDEDNEKEVSVVNTLFSKQVDGIIYMGYHLTDKIRSEFSRSRTPIVLAGTVDVEHQLPSVNIDYK
QATIDAVSYLAKENERIAFVSGPLVDDINGKVRLVGYKETLKKAGITYSEGLVFESKYSYDDGYALAERLISSNATAAVVTGD
ELAAGVLNGLADKGVSVPEDFEIITSDDSQISRFTRPNLTTIAQPLYDLGAISMRMLTKIMHKEELEEREVLLPHGLTERSST
RKRK


SeqID 226
MKKKLVFPNLFWWGAASSGPQTEGQYGKVHENVMDYWFKTHPEDFFDNVGPLVASNFFHTYTEDFHLMKEIGVNSFRTSIQWS
RLIKNLETGEPDPKGIAFYNAIIEEAKKNQMDLVMNLHHFDLPVELLQKYGGWESKHVVELFVKFAKTAFTCFGDKVHYWTTF
NEPMVIPEAGYLYAFHYPNLKGKGKEAVQVIYNLNLASAKVIQLYRSLELDGKIGIILNLTPAYPRSNSPEDLEASRFTDDFF
NKVFLNPAVKGTFPERLVKQLERDGVLWSHTEKELQLMKSNTVDFLGVNYYHPKRVQAQANPEEYQTPWMPDQYFKEYEWLER
RMNPYRGWEIFPKAIYDIAMIVKEBYGNIPWFISENGMGVENEARFIDENGVIDDVYRIEFYEEHLRWLHKAIEEGSHCFGYH
AWTAFDCWSWNNAYKNRYGFISVDLETQKRTIKSSGRWYRKVSDNNGFEVEIEE


SeqID 227
VENLTNFYEKYRVYLTRPRLELLAVVTIVFCAVLVFFLNIPGKGVLKLDNGTIVYDGSLVRGKMNGQGTITFQNGDQYTGGFN
NGAFNGKGTFQSKEGWTYEGDFVNGQAEGKGKLTTEQEVVYEGTFKQGVFQQK


SeqID 228
MLNKIRDYLDFAGLQYRNPDKAGAEREKMLAFRHKGQEARKVFTELAKAFQASHPEWQLQQTSQWMNQAQRLRPHFWVYLQRD
GQVTEPMMALRLYGTSTDFGISLEVSFIERKKDEQTLGKQAKVLDIPTVKGIYYLTYSNGQSQRWEANEEKRRTLREKVRSQE
VRKVLVKVDVPMTENSSEEEIVEGLLKSYSKILPYYLATRK


SeqID 229
MVQNSCWQSKSHKVKAFTLLESLLALIVISGGLLLFQAMSQLLISEVRYQQQSEQKEWLLFVDQLEVELDRSQFEKVEGNRLY
MKQDGKDIAIGKSKSDDFRKTNARGRGYQPMVYGLKSVRITEDNQLVRFHFQFQKGLEREFIYRVEKEKS


SeqID 230
MKKMMTFLKKAKVKAFTLVEMLVVLLIISVLFLLFVPNLTKQKEAVNDKGKAAVVKVVESQAELYSLEKNEDASLRKLQADGR
ITEEQAKAYKEYNDKNGGANRKVND


SeqID 231
```

MTSKVRKAVIPAAGLGTRFLPATKALAKEMLPIVDKPTIQFIVEEALKSGIEDILVVTGKSKRSIEDHFDSNFELEYNLKEKG
KTDLLKLVDKTTDMRLHFIRQTHPRGLGDAVLQAKAFVGNEPFVVMLGDDLMDITDEKAVPLTKQLMDDYERTHASTIAVMPV
PHDEVSAYGVIAPQGEGKDGLYSVETFVEKPAPEDAPSDLAIIGRYLLTPEIFEILEKQAPGAGNEIQLTDAIDTLNKTQRVF
AREFKGARYDVGDKFGFMKTSIDYALKHPQVKDDLKNYLIQLGKELTEKE

SeqID 232

MQNQLNELKRKMLEFFQQKQKNKKSARPGKKGSSTKKSKTLDKSVIFPAILLSIKALFNLLFVLGFLGGMLGAGIALGYGVAI
FDKVRVPQTEELVNQVKDISSISEITYSDGTVIASIESDLLRTSISSEQISENLKKAIIATEDEHFKEHGVVPKAVIRATLG
KFVGLGSSSGGSTLTQQLIKQQVVGDAPTLARKAAEIVDALALERAMNKDEILTTYLNVAPFGRNNKGQNIAGARQAAEGIFG
VDASQLTVPQAAFLAGLPQSPITYSPYENTGELKSDEDLEIGLRRAKAVLYSMYRTGALSKDEYSQYKDYDLKQDFLPSGTVI
GISRDYLYFTTLAEAQERMYDYLAQRDNVSAKELKNEATQKFYRDLAAKEIENGGYKITTTIDQKIHSAMQSAVADYGYLLDI
GTGRVEVGNVLMDNQTGAILGFVGGRNYQENQNNHAFDTKRSPASTTKPLLAYGIAIDQGLMGSETILSNYPTNFANGNPIMY
ANSKGTGMMTLGEALNYSWNIPAYWTYRMLREKGVDVKGYMEKMGYEIPEYGIESLPMGGGIEVTVAQHTNGYQTLANNGVYF
QKHVISKIEAADGRVVYEYQDKPVQVYSKATATIMQGLLREVLSSRVTTTFKSNLTSLNPTLANADWIGKTGTTNQDENMWLM
LSTPRLTLGGWIGHDDNHSLSRRAGYSNNSNYMAHLVNAIQQASPSIWGNERFALDPSVVKSEVLKSTGQKPEKVSVEGKEVE
VTGSTVTSYWANKSGAPATSYRFAIGGSDADYQNAWSSIVGSLPTPSSSSSSSSSSSSDSSNSSTTRPSSSRARR

SeqID 233

MSSKFMKSAAVLGTATLASLLLVACGSKTADKPADSGSSEVKELTVYVDEGYKSYIEEVAKAYEKEAGVKVTLKTGDALGGLI
KLSLDNQSGNVPDVMMAPYDRVGSLGSDGQLSEVKLSDGAKTDDTTKSLVTAANGKVYGAPAVIESLVMYYNKDLVKDAPKTF
ADLENLAKDSKYAFAGEDGKTTAFLADWTNFYYTYGLLAGNGAYVFGQNGKDAKDIGLANDGSIVGINYAKSWYEKWPKGMQI
TEGAGNLIQTQFQEGKTAAIIDGPWKAQAFKDAKVNYGVATIPTLPNGKEYAAFGGGKAWVIPQAVKNLEASQKFVDFLVATF
QQKVLYDKTNEIPANTEARSYAEGKNDELTTAVIKQFKNTQPLPNISQMSAVWDPAKNMLFDAVSGQKDAKTAANDAVTLIKF
TIKQKFGE

SeqID 234

MIDKVVRNLLLTFFFCKMTKIIIFLTTILVKKKKICYNEFKLRNRKQKGVIMWVLGFILFMIFFYSNNSKKIKKLENKIKRLF
RKEKGNAEMSRLLQEMIGKEPIITGVYIGPDNWEVVDVDEEWVKLRRVDNTGKEKFKLQRIEDIQTVEFDGE

SeqID 235

MILSKNREDGLRKFATNIRLNTLRTLNHLGFGHYGGSLSIVEVLAVLYGEIMPMTPEIFAARDRDYFILSKGHGGPALYSTLY
LNGFFDKEFLYSLNTNGTKLPSHPDRNLTPGIDMTTGSLGQGISVATGLAYGQRIRKSPFYTYAIVGDGELNEGQCWEAIQFA
SHQQLSNLIVFVDDNKKQLDGFTKDICNPGDFVEKFSAFGFESIRVKGSDIREIYEGIVQLKQSNNSSPKCIVLDTIKGQGVG
ELEEMKSNHHLRPTVEEKQMLTSVVERLSQELEETE

SeqID 236

MKKTTILSLTTAAVILAAYVPNEPILADTPSSEVIKETKVGSIIQQNNIKYKVLTVEGNIGTVQVGNGVTPVEFEAGQDGKPI
TIPTKITVGDKVFTVTEVASQAFSYYPDETGRIVYYPSSITIPSSIKKIQKKGFHGSKAKTIIFDKGSQLEKIEDRAFDFSEI
EEIELPASLEYIGTSAFSFSQKLKKLTFSSSSKLELISHEAFANLSNLEKLTLPKSVKTLGSNLFRLTTSLKHVDVEEGNESI
ASVDGVLFSKDKTQLIYYPSQKNDESYKTPKETKELASYSFNKNSYLKKLELNEGLEKIGTFAFADAIKLEEISLPNSLETII
RLAFYGNLELKELILPDNVKNFGKHVMNGLPKLKSLTIGNNINSLPSFFLSGVLDSLKEIHIKNKSTEFSVKKDTFAIPETVI
FYVTSEHIKDVLKSNLSTSNDIIVEKVDNIKQETDVAKPKKNSNQGVVGWVKDKGLWYYLNESGSMATGWVKDKGLWYYLNES
GSMATGWVKDKGLWYYLNESGSMATGWVKDKGLWYYLNESGSMATGWVKDKGLWYYLNESGSMATGWVKDKGLWYYLNESGSM
ATGWVTVSGKWYYTYNSGDLLVNTTTPDGYRVNANGEWVG

SeqID 237

MVRFTGLSLKQTQAIEVLKGHISLPDVEVAVTQSDQASISIEGEEGHYQLTYRKPHQLYRALSLLVTVLAEADKVEIEEQAAY
EDLAYMVDCSRNAVLNVASAKQMIEILALMGYSTFELYMEDTYQIEGQPYFGYFRGAYSAEELQEIEAYAQQFDVTFVPCIQT
LAHLSAFVKWGVKEVQELRDVEDILLIGEEKVYDLIDGMFATLSKLKTRKVNIGMDEAHLVGLGRYLILNGVVDRSLLMCQHI
ERVLDIADKYGFHCQMWSDMFFKLMSADGQYDRDVEIPEETRVYLDRLKDRVTLVYWDYYQDSEEKYNRNFRNHHKISHDLAI
AGGAWKWIGFTPHNHFSRLVAIEANKACRANQIKEVIVTGWGDNGGETAQFSILPSLQIWAELSYRNDLDGLSAHFKTNTGLY
VEDFMQIDLANLLPDLPGNLSGINPNRYVFYQDILCPILDQHMTPEQDKPHFAQAAETLANIKEKAGNYAYLFETQAQLNAII
SSKVDVGRRIRQAYQADDKESLQQIARQELPELRSQIEDFHALFSHQWLKENKVFGLDTVDIRMGGLLQRIKRAESRIEVYLA
GQLDRIDELEVEILPFTDFYADKDFAATTANQWHTIATASTIYTT

SeqID 238

MSNSFVKLLVSQLFANLADIFFRVTIIANIYIISKSVIATSLVPILIGISSFVASLLVPLVTKRLALNRVLSLSQFGKTILLA
ILVGMFTVMQSVAPLVTYLFVVAISILDGFAAPVSYAIVPRYATDLGKANSALSMTGEAVQLIGWGLGGLLFATIGLLPTTCI
NLVLYIISSFLMLFLPNAEVEVLESETNLEILLKGWKLVARNPRLRLFVSANLLEIFSNTIWVSSIILVFVTELLNKTESYWG
YSNTAYSIGIIISGLIAFRLSEKFLAAKWEPQLFTPNLKTIQNPCLSLDPGWFLFSPNGCFLLDKKEFPLYGISVEKNTKRKI
THMNSLPNHHFQNKSFYQLSFDGGHLTQYGGLIFFQELFSQLKLKERISKYLVTNDQRRYCRYSDSDILVQFLFQLLTGYGTI
YACKELSADAYFPKLLEGGQLASQPTLSRFLSRTDEETVHSLRCLNLELVEFFLQFHQLNQLIVDIDSTHFTTYGKQEGVAYP
AHYRAHGYHPLYAFEGKTGYCFNAQLRPGNRYCSEEADSFITPVLERFNQLLFRMDSGFATPKLYDLIEKTGQYYLIKLKKNT
VLSRLGDLSLPCPQDEDLTILPHSAYSETLYQAGSWSHKRRVCQFSERKEGNLFYDVISLVTNMTSGTSQDQFQLYRGRGQAI
NPIKEMKEGPFGDKTDSSTLIKNEVRMMMSCIAYNLYLFLKHLAGGDFQTLTIKRFRHLFLHVVGKCVRTGRKQLLKLSSLYA
YSELFSALYSRIRKVNLNLPVPYEPPRRKASLMMH

SeqID 239

MFASKSERKVHYSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQGEQPKKLDSERDKARKEVEEYVK
KIVGESYAKSTKKRHTITVALVNELNNIKNEYLNKIVESTSESQLQILMMESRSKVDEAVSKFEKDSSSSSSSDSSTKPEASD
TAKPNKPTEPGEKVAEAKKKVEEAEKKAKDQKEEDRRNYPTITYKTLELEIAESDVEVKKAELELVKVKANEPRDEQKIKQAE
AEVESKQAEATRLKKIKTDREEAEEEAKRRADAKEQGKPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKK
VAEAEKKVEEAKKKAEDQKEEDRRNYPTNTYKTLELEIAESDVEVKKAELELVKEEAKEPRNEEKVKQAKAEVESKKAEATRL
EKIKTDRKKAEEEAKRKAAEEDKVKEKPAEQPQPAPAPKAEKPAPAPKPENPAEQPKAEKPADQQAEEDYARRSEEEYNRLTQ
QQPPKTEKPAQPSTPKTGWKQENGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATGWLQNNGSWYYLNANGSMATGWLQNN
GSWYYLNANGSMATGWLQYNGSWYYLNANGSMATGWLQYNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMKASQWFKVSDK
WYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 240
MNYSKALNECIESAYMVAGHFGARYLESWHLLIAMSNHSYSVAGATLNDYPYEMDRLEEVALELTETDYSQDETFTELPFSRF
LQVLFDEAEYVASVVHAKVLGTEHVLYAILHDSNALATRILERAGFSYEDKKDQVKIAALRRNLEERAGWTREDLKALRQRHF
TVADKQNSMANMMGMPQTPSGGLEDYTHDLTEQARSGKLEPVIGRDKEISRMIQILSRKTKNNPVLVGDAGVGKTALALGLAQ
RIASGDVPAEMAKMRVLELDLMNVVAGTRFRGDFEERMNNIIKDIEEDGQVILFIDELHTIMGSGSGIDSTLDAANILKPALA
RGTLRTVGATTQEEYQKHIEKDAALSRRFAKVTIEEPSVADSMTILQGLKATYEKHHRVQITDEAVETAVKMAHRYLTSRHLE
DSAIDLLDEAAATVQNKAKHVKADDSDLSPADKALMDGKWKQAAQLIAKEEEVPVYKDLVTESDILTTLSRLSGIPVQKLTQI
DAKKYLNLEAELHKRVIGQDQAVSSISRAIRRNQSGIRSHKRPIGSFMFLGPTGVGKTELAKALAEVLFDDESALIRFDMSEY
MEKFAASRLNGAPPGYVGYEEGGELTEKVRNKPYSVLLFDEVEKAHPDIFNVLLQVLDDGVLTDSKGRKVDFSNTIIIMTSNI
GATALRDDKTVGFGAKDIRFDQENMEKRMFEELKKAYRPEFINRIDEKVVFHSLSSDHMQEVVKIMVKPLVASLTEKGIDLKI
QASALKLLANQGYDPEMGARPLRRTLQTEVEDKLAELLLKGDLVAGSTLKIGVKAGQLKFDIA

SeqID 241
MKILPFIARGTSYYLKMSVKKLVPFLVVGLMLAAGDSVYAYSRGNGSIARGDDYPAYYKNGSQEIDQWRMYSRQCTSFVAFRI
SNVNGFEIPAAYGNANEWGHRARREGYRVDNTPTIGSITWSTAGTYGHVAWVSNVMGDQIEIEEYNYGYTESYNKRVIKANTN
TGFIHFKDLDGGSVGNSQSSTSTGGTHYFKTKSAIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSYRYVQI
EAVNKNPLGNSVLSSTGGTHYFKTKSAIKTEPLVSATVIDYYYPGEKVHYDQILEKDGYKWLSYTAYNGSRRYIQLEGVTSSQ
NYQNQSGNISSYGSHSSSTVGWKKINGSWYHFKSNGSKSTGWLKDGSSWYYLKLSGEMQTGWLKENGLWYYLGSSGAMKTGWY
QVSGKWYYSYSSGALAVNTTVDGYRVNSDGERV

SeqID 242
MKVIFLADVKGKGKKGEIKEVPTGYAQNFLIKKNLAKEATAQAVGELRGKQKSEEKAHAEMIAEGKAIKAQLEAEETVVEFVI
KVGPDGRTFGSITNKKIAEELQKQFGIKIDKRHIQVQAPIRAVGLIDVPVKIYQDITSVINLRVKEG

SeqID 243
MKKKILASLLLSTVMVSQVAVLTTAHAETTDDKIAAQDNKISNLTAQQQEAQKQVDQIQEQVSAIQAEQSNLQAENDRLQAES
KKLEGEITELSKNIVSRNQSLEKQARSAQTNGAVTSYINTIVNSKSITEAISRVAAMSEIVSANNKMLEQQKADKKAISEKQV
ANNDAINTVIANQQKLADDAQALTTKQAELKAAELSLAAEKATAEGEKASLLEQKAAAEAEARAAAVAEAAYKEKRASQQQSV
LASANTNLTAQVQAVSESAAAPVRAKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAQWATSAAAAGFRTGSTPQVGA
IACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTTTSEGFVTYIYAD

SeqID 244
MVKRRIRRGTREPEKVVVPEQSSIPSYPVSVTSNQGTDVAVEPAKAVAPTTDWKQENGMWYFYNTDGSMATGWVQVNSSWYYI
NSNGSMKVNQWFQVGGKWYYVNTSGELAVNTSIDGYRVNDNGEWVR

SeqID 245
ELRRLSRLVDQELYFGCGWRLSLEWLPSMRKDSWPSNTAPRTTMVQ

SeqID 246
DCIRKQPFTRDEPNKTCRKTKPSKSYCSYRW

SeqID 247
GQRNPRRIERVIRMAETKPRISKKEG

SeqID 248
QRKLFKIFHLFQKKSGWNQKSSCLKLNLNSLNRKMTQMTKMFRSIFQPKKPLNTNFQAYNSLHQINQKISLKRRKLSEKISKS

SeqID 249
LVIIVLKIQSKSETDFIFKTWPFILLSKIIPLMVLDCQVSISWTNRETVAYSKLLAIKTLKGDYHDGQSKKIRLSHASRVRTI
SWYPHDMADSTRIMAFSRKGC

SeqID 250
ERLPAFPRSLSGRKLDQGGTKEKGSDGRSP

SeqID 251
RNCLSTWKSSSNYHTEIKRGTVRQCLGKGRFKEVYSADYAQQSYENNRKRSVKKSSLTKELKEKILHYHNQKFSPEMMVMAKQ
VNVGISTIYYWIHHGKLGLSKQDLLYPRKGKALKKQASTNFKPAGQSIEQRPEAINLRLENGHYEIDTVLLTRSKNYCLIVLN
DRKSRHQIIRLIPNKSAEVVNQALKLILKQHKILSITADNGTEFNRLFDIFSEEHIYYAHPYASWERGTNENHNRLIRR

SeqID 252
PVMTISSPTMKNMDLSTKASPSQPLQGKHGMIWSGK

SeqID 253
TSSIRIHTRKSSPNWTTTPHLALSAETN

SeqID 254
YFLPHKYARESLSLPSTNKILHRKQGS

SeqID 255
AAFKKDQINERVEKLGKLKPITINYNGKSEVIDSKEKLQELMNKAVKDEVAQI

SeqID 256
AYAHSKRSAGSGRAGGRQCLCQCQNKCRRDFKY

SeqID 257
HGRPYHKPHQPHHHGFPQQSYNLLPPKHKPTLCVRR

SeqID 258
KGKILLLPRLTTQRWQRKIRPDSRKSANNKANLDFHNSRCKSSLSDEAPNLHKNPALLVHSLSRVIAVLLELSPLQAYSIVKI
SPKEDDLIHDDAILVRFGILEVHDSPYELLLLYHTHSYRFSCSIYLS

SeqID 259
FTVSHVFLLYLSFNPRPKSMSLSFTSSKLLRPRFRTFIISASDFSVKSCTVLIPARFKQLYERTDKSSSSIVRSKIRSSDSV:
ASFITSVDLAISVRFVNKSRCSVKIRAESPKASSGIIVPFVKISRVNLSKPSLLPTRAGST

SeqID 260
LVCMKNKGCYKERNNCCH

SeqID 261
FHYLSKYFLVSAITTGDKTKRAIKFGRAIKALTISAIIQTISNSINPPNKTMSTKTTR

SeqID 262
GKKVFIKYPLSRVSSKTGPMITGRTKDKIVDKKVGCPFEKSTVKYSS

SeqID 263
SSPVFPKLVMVSGANKPRERRNFPFSSKMSFHLTFVL

SeqID 264
YLTSFSVPKIASSKVKLTRYWRSSPWRGAFGLREEPPPPKKLEKISSKPPKPPAPLKPPKPPAPPKPPLAPAAPYWS

SeqID 265
QSWRPIPDSKCYTQEKLTIPIKRRKDIKDFYHNSIQRHKNSHKSHLLDSYRLIITRLAEIVHENKILIVLILYVTNIPSRSRI
YEVNRV

SeqID 266
LFRFYRVIVLYRGWHIYLLILVNLQYVQNVFRKDRFLVRGAQPFFHGERSAGHLVLPYVL

SeqID 267
ITHPPLNPEHFVSRVFSSLGLKSYQPKDDRFLRKPSDSRHPESGNSGKWQVLNSPLVIVK

SeqID 268
TLAKAVGLMYSPPIPPKPFLGRITTDSSSI

SeqID 269
PGSPFSEISGAGFFGVAKRIFPRPPRPPWATINSCPCSIKSVKTLPVSASRTVVPCGTRTLRSSAPRPCIPLVIPFSPESAL!

SeqID 270
INSLTLATSLSKRRAPRKASKASPRMVSRLRPPDFSSPLPNLINWSNWQSRAKPAKLSSRTIIARSFDRSPSGFLGYFLYRY:
EINNCRTASPKNSKRSLCEIFKRRCSLA

SeqID 271
PIGKRNCKAECQSHHLLEKQKTFQSRKTKRYGASPEPRYRESRKPRLSQ

SeqID 272
QPLGHSKAEEHETICSHTFDNHTTETIPNQVKGRDMTSSETLPFPSKNQNQGKAKQIP

SeqID 273
PCSLPDYGLVGSGYHSCHYQSNDTRFLESHGNWRTLLYSWSWILCQEKTLFPHDLASLYPSCVRTSIHRYCLLHVKKLRNSIS
TFFFTHIDKVLVQAHIISQFWMKRTYQHIFFLGCNNLIVHCC

SeqID 274
RVKDNHLDKLVKALLKRRSSTQYYIHQLLRKMIRLYGHQQLHNNSEILVYSDYGHVDLLLLETNKIPVY

SeqID 275
QVIKIDIATTNKTESVKSQSEREKKRLTSSNILKVRGRPI

SeqID 276
AFKSSKVPSLDPSSTKTYSISVSKSGSKASRA

SeqID 277
DKTDPLARKLPDKSKPSTSFCTKSLSPVNMASLTSAKPSKTKASLGICSPADKRMTSPSTNSSGLRATS

SeqID 278
LCRLQTQARPRGSVTNLTKQNKVYRYLNYLRQTQLSAM

SeqID 279
CIQSIGNEGQCKGNSCYVGKEIHLAPISDIVGHKGKEEGDDGNDDGRQFYLFLAHLVGSAFLRSFPLLYSKGINEEGDGIGND
GRLISKDIINSTGQNGINNTKVISPFA

SeqID 280
TSTKLVIDTTTFMTFCTNNTKSSKFTNTFTKLDVGTTTRHVGCDGNGTTLTSIHDDLGFSIVVFGIQDFVRNTSCNQFLRNVV
TSFNRYCTNQDRLTLLVTSLNVFDNRFKLRFDTCIKKV

SeqID 281
ISTGVPTCSMYPLLKTAILSDKVKASS

SeqID 282
LSTMSSIKSILCSLPTHTIFSKSAISPKRGSTAI

SeqID 283
SCSYSSTNSKACSIVGLSKPTIFICVTPISAAKAISSASRPANSSTFKCVCVSKIISSNLVFYLLLYQKRGGAPNFSVSPPLF
NRELFCYLFYPILPISYSTVRDRRDWLH

SeqID 284
VTSCIVPAVACGALVVLGAALGATGLLGTVTMAMACTPIVSASFTSSIVFAFSISLRAACFSASTLAKSSAFSLSESGAPLIS
SCLSLAAFSMAFLADSFSVANCLEACD

SeqID 285
YSPFNHSILIRKTTKIINPNPKAPRMNWRSKVWSNQPVNISTNHTKSDRPIKK

SeqID 286
DYFKFRTTFTRFSTVKPYSANTFGAGAEAPKVSIPRTAPSRPTYLYQF

SeqID 287
RGRRGLCVARIKAPRLVIKPKRTIDPPTKDRYSPPLSATSLSVPKSPIISFPAKMDKNPKRKLNSKVIFNASVT

SeqID 288
TPPYTKIPAKTAIIPFISAQDFNQAQRLSG

SeqID 289
ATGAAGAAAAAAATCTTAGCGTCACTTTTATTAAGTACAGTAATGGTTTCTCAAGTAGCTGTTTTAACAACTGCGCATGCAGA
AACGACTGATGACAAAATTGCTGCTCAAGATAATAAAATTAGTAACTTAACAGCACAACAACAAGAAGCCCAAAAACAAGTTG
ACCAAATTCAGGAGCAAGTATCAGCTATTCAAGCTGAGCAGTCTAACTTGCAAGCTGAAAATGATAGATTACAAGCAGAATCT
AAGAAACTCGAGGGTGAGATTACAGAACTTTCTAAAAACATTGTTTCTCGTAACCAATCGTTGGAAAAACAAGCTCGTAGTGC
TCAAACAAATGGAGCCGTAACTAGCTATATCAATACCATTGTAAACTCAAATCAATTACAGAAGCTATTTCACGTGTTGCTG
CAATGAGTGAAATCGTATCTGCAAACAACAAAATGTTAGAACAACAAAAGGCAGATAAAAAAGCTATTTCTGAAAAACAAGTA
GCAAATAATGATGCTATCAATACTGTAATTGCTAATCAACAAAAATTGGCTGATGATGCTCAAGCATTGACTACGAAACAGGC
AGAACTAAAAGCTGCTGAATTAAGTCTTGCTGCTGAGAAAGCGACAGCTGAAGGGGAAAAAGCAAGTCTATTAGAGCAAAAAG
CAGCAGCTGAGGCAGAGGCTCGTGCAGCTGCGGTAGCAGAAGCAGCTTATAAAGAAAAACGAGCTAGCCAACAACAATCAGTA
CTTGCTTCAGCAAACACTAACTTAACAGCTCAAGTGCAAGCAGTATCTGAATCTGCAGCACCTGTCCGTGCAAAAGTTCG
TCCAACATACAGTACAAACGCTTCAAGTTATCCAATTGGAGAATGTACATGGGGAGTAAAAACATTGGCACCTTGGGCTGGAG
ACTACTGGGGTAATGGAGCACAGTGGGCTACAAGTGCAGCAGCAGCAGGTTTCCGTACAGGTTCAACACCTCAAGTTGGAGCA
ATTGCATGTTGGAATGATGGTGGATATGGTCACGTAGCGGTTGTTACAGCTGTTGAATCAACAACACGTATCCAAGTATCAGA

ATCAAATTATGCAGGTAATCGTACAATTGGAAATCACCGTGGATGGTTCAATCCAACAACAACTCCTGAAGGTTTTGTTACAT
ATATTTATGCAGAT

SeqID 290
ATGAAGAAAAAAATCTTAGCGTCACTTTTATTAAGTACAGTAATGGTTTCTCAAGTAGCTGTTTTAACAACTGCGCATGCAGA
AACGACTGATGACAAAATTGCTGCTCAAGATAATAAAATTAGTAACTTAACAGCACAACAACAAGAAGCCCAAAAACAAGTTG
ACCAAATTCAGGAGCAAGTATCAGCTATTCAAGCTGAGCAGTCAACTTGCAAGCTGAAAATGATAGATTACAAGCAGAATCT
AAGAAACTCGAGGGTGAGATTACAGAACTTTCTAAAAACATTGTTTCTCGTAACCAATCGTTGGAAAAACAAGCTCGTAGTGC
TCAAACAAATGGAGCCGTAACTAGCTATATCAATACCATTGTAAACTCAAAATCAATTACAGAAGCTATTTCACGTGTTGCTG
CAATGAGTGAAATCGTATCTGCAAACAACAAAATGTTAGAACAACAAAAGGCAGATAAAAAAGCTATTTCTGAAAAACAAGTA
GCAAATAATGATGCTATCAATACTGTAATTGCTAATCAACAAAAATTGGCTGATGATGCTCAAGCATTGACTACGAAACAGGC
AGAACTAAAAGCTGCTGAATTAAGTCTTGCTGCTGAGAAAGCGACAGCTGAAGGGGAAAAAGCAAGTCTATTAGAGCAAAAAG
CAGCAGCTGAGGCAGAGGCTCGTGCAGCTGCGGTAGCAGAAGCAGCTTATAAAGAAAAACGAGCTGGCCAACAACAATCAGTA
CTTGCTTCAGCAAACACTAACTTAACAGCTCAAGTGCAAGCAGTATCTGAATCTGCAGCAGCACCTGTCCGTGCAAAAGTTCG
TCCAACATACAGTACAAACGCTTCAAGTTATCCAATTGGAGAATGTACATGGGGAGTAAAAACATTGGCACCTTGGGCTGGAG
ACTACTGGGGTAATGGACGACAGTGGGCTACAAGTGCAGCAGCAGCAGGTTTCCGTACAGGTTCAACACCTCAAGTTGGAGCA
ATTGCATGTTGGAATGATGGTGGATATGGTCACGTAGCGGTTGTTACAGCTGTTGAATCAACAACACGTATCCAAGTATCAGA
ATCAAATTATGCAGGTAATCGTACAATTGGAAATCACCGTGGATGGTTCAATCCAACAACAACTTCTGAAGGTTTTGTTACAT
ATATTTATGCAGAT

SeqID 291
ATGATCCAAATCGGCAAGATTTTTGCCGGACGCTATCGGATTGTCAAACAGATTGGTCGAGGAGGTATGGCGGATGTCTACCT
AGCCAAAGACTTAATCTTAGATGGGGAAGAAGTGGCAGTGAAGGTTCTGAGGACCAACTACCAGACGGACCCGATAGCTGTAC
CTCGTTTTCAGCGTGAAGCGAGAGCTATGGCAGATCTAGACCATCCTCATATCGTTCGGATAACAGATATTGGCGAGGAAGAC
GGTCAACAGTACCTAGCTATGGAGTATGTGGCTGGACTGGACCTCAAACGCTATATCAAGGAACATTATCCTCTTTCTAATGA
AGAAGCAGCCCGTATCATGGGACAAATTCTCTTGGCTATGCGCTTGGCCCATACTCGAGGAATTGTTCACAGGGACTTGAAAC
CTCAAAATATCCTCTTGACACCAGATGGGACTGCCAAGGTCACAGACTTTGGGATTGCTGTAGCCTTTGCAGAGACAAGCTCTG
ACCCAGACTAACTCGATGTTGGGCTCAGTTCATTACTTGTCACCAGAGCAGGCGGGTGGTTCGAAGGCGACTGTGCAGAGTGA
TATCTATGCCATGGGGATTATTTTCTATGAGATGCTGACAGGCCATATCCCTTATGACGGGGATAGCGCGGTGACCATTGCCC
TCCAGCATTTCCAGAAACCCCTGCCGTCCGTTATTGCAGAAAATCCATCTGTACCTCAGGCTTTAGAAAATGTTATTATCAAG
GCAACTGCTAAAAAGTTGACCAATCGCTACCGCTCGGTTTCAGAGATGTATGTGGACTTGTCTAGTAGCTTGTCCTACAATCG
TAGAAATGAAAGTAAGTTAATCTTTGATGAAACGAGCAAGGCAGATACCAAGACCTTGCCGAAGGTTTCTCAGAGTACCTTGA
CATCTATTCCTAAGGTTCAAGCGCAAACAGAACACAAATCAATCAAAAACCCAAGCCAGGCTGTGACAGAGGAAACTTACCAA
CCACAAGCACCGAAAAAACATAGATTTAAGATGCGTTACCTGATTTTGTTGGCCAGCCTTGTATTGGTGGCAGCTTCTCTTAT
TTGGATACTATCCAGAACTCCTGCAACCATTGCCATTCCAGATGTGGCAGGTCAGACAGTTGCAGAGGCCAAGGCAACGCTCA
AAAAAGCCAATTTTGAGATTGGTGAGGAGAAGACAGAGGCTAGTGAAAAGGTGGAAGAAGGGCGGATTATCCGTACAGATCCT
GGCGCTGGAACTGGTCGAAAAGAAGGAACGAAAATCAATTTGGTTGTCTCATCAGGCAAGCAATCTTTCCAAATTAGTAATTA
TGTCGGTCGGAAATCCTCTGATGTCATTGCGGAATTAAAAGAGAAAAAAGTTCCAGATAATTTGATTAAAAATTGAGGAAGAAC
AGTCGAATGAGAGTGAGGCTGGAACGGTCCTGAAGCAAAGTCTACCAGAAGGTACGACCTATGACTTGAGCAAGGCAACTCAA
ATTGTCTTGACAGTAGCTAAAAAAGTTACAAGTGTTGCCATGCCGAGTTACATTGGTTCTAGCTTGGAGTTTACTAAGAACAA
TTTGATTCAAATTGTTGGGATTAAGGAAGCTAATATAGAAGTTGTAGAAGTGACGACAGCGCCTGCAGGTAGTGCAGAAGGCA
TGGTTGTTGAACAAAGTCCTAGAGCAGGTGAAAAGGTAGACCTCAATAAGACTAGAGTCAAGATTTCAATCTACAAACCTAAA
ACAACTTCAGCTACTCCT

SeqID 292
ATGATCCAAATCGGCAAGATTTTTGCCGGACGCTATCGGATTGTCAAACAGATTGGTCGAGGAGGCATGGCGGATGTCTACCT
AGCCAAAGACTTAATCTTAGATGGGGAAGAAGTGGCAGTGAAGGTTCTGAGGACCAACTACCAGACGGACCCGATAGCTGTAC
CTCGTTTTCAGCGTGAAGCGAGAGCTATGGCAGATCTAGACCATCCTCATATCGTTCGGATAACAGATATTGGTGAGGAAGAC
GGTCAACAGTATCTTGCAATGGAGTATGTTGCTGGACTAGACCTCAAACGCTATATCAAGGAACATTATCCTCTTTCTAATGA
AGAAGCAGTCCGTATCATGGGACAAATTCTCTTGGCTATGCGCTTGGCCCATACTCGAGGAATTGTTCACAGGGACTTGAAAC
CTCAAAATATCCTTTTGACACCAGATGGGACGGCCAAGGTCACAGACTTTGGGATTGCTGTAGCCTTTGCAGAGACAAGCTCTG
ACCCAGACTAACTCGATGTTGGGCTCAGTTCATTACTTGTCACCAGAGCAGGCGCGTGGTTCGAAGGCGACTGTGCAGAGTGA
TATCTATGCCATGGGGATTATTTTCTATGAGATGTTGACAGGCCATATCCCTTATGACGGGGATAGCGCGGTGACCATTGCCC
TCCAGCATTTCCAGAAACCCCTGCCGTCCGTTATTGCAGAAAATCCATCTGTACCTCAGGCTTTAGAAAATGTTATTATCAAG
GCAACTGCTAAAAAGTTGACCAATCGCTACCGCTCGGTTTCAGAGATGTATGTGGACTTGTCTAGTAGCTTGTCCTACAATCG
TAGAAATGAAAGTAAGTTAATCTTTGATGAAACGAGCAAGGCAGATACCAAGACCTTGCCGAAGGTTTCTCAGAGTACCTTGA
CATCTATTCCTAAGGTTCAAGCGCAGACAGAACACAAATCAATCAAAAACCCAAGCCAGGCTGTGACAGAGGAAACTTACCAA
CCACAAGCACCGAAAAAACATAGATTTAAGATGCGTTACCTGATTTTGTTGGCCAGCCTTGTATTGGTGGCAGCTTCTCTTAT
TTGGATACTATCCAGAACTCCTGCAACCATTGCCATTCCAGATGTGGCAGGTCAGACAGTTGCAGAGGCCAAGGCAACGCTCA
AAAAAGCCAATTTTGAGATTGGTGAGGAGAAGACAGAGGCTAGTGAAAAGGTGGAAGAAGGGCGGATTATCCGTACAGATCCT
GGCGCTGGAACTGGTCGAAAAGAAGGAACGAAAATCAATTTGGTTGTCTCATCAGGCAAGCAATCTTTCCAAATTAGTAATTA
TGTCGGTCCGGAAATCTTCTGATGTTATCGCGGAATTAAAAGAGAAAAAAGTTCCAGATAATTTGATTAAAAATTGAGGAAGAAC
AGTCGAATGAGAGTGAGGCTGGAACGGTCCTGAAGCAAAGTCTACCAGAAGGTACGACCTATGACTTGAGCAAGGCAACTCAA
ATTGTTTTGACAGTAGCTAAAAAAGCTACGACGATTCAATTAGGGAACTATATTGGACGGAACTCTACAGAAGTAATCTCAGA
ACTCAAGCAGAAGAAGGTTCCTGAGAATTTGATTAAGATAGAGGAAGAAGAGTCCAGCGAAAGCGAACCAGGAACGATTATGA
AACAAAGTCCAGGTGCCGGAACGACTTATGATGTGAGTAAACCTACTCAAATTGTCTTGACAGTAGCTAAAAAAGTTACAAGT
GTTGCCATGCCGAGTTACATTGGTTCCAGCTTGGAGTTTACTAAGAACAATTTGATTCAAATTGTTGGGATTAAGGAAGCTAA

TATAGAAGTTGTAGAAGTGACGACAGCGCCTGCAGGTAGTGTAGAAGGCATGGTTGTTGAACAAAGTCCTAGAGCAGGTGAAA
AGGTAGACCTAAATAAGACTAGAGTCAAGATTTCAATCTACAAACCTAAAACAACTTCAGCTACTCCT

SeqID 293
ATGATCCAAATCGGCAAGATTTTTGCCGGACGCTATCGGATTGTCAAACAGATTGGTCGAGGAGGTATGGCGGATGTCTACCT
AGCCAAAGACTTAATCTTAGATGGGGAAGAAGTGGCAGTGAAGGTTCTGAGGACCAACTACCAGACGGACCCGATAGCTGTAG
CTCGTTTTCAGCGTGAAGCGAGAGCTATGGCAGATCTAGACCATCCTCATATCGTTCGGATAACAGATATTGGTGAGGAAGAC
GGTCAACAGTATCTTGCAATGGAGTATGTTGCTGGACTAGACCTCAAACGCTATATCAAGGAACATTATCCTCTTTCTAATGA
AGAAGCAGTCCGTATCATGGGACAAATCCTCCTAGCCATGCGTGTTTGGCCCATACCAGAGGAATTGTTCACAGGGACTTGAAAC
CTCAAAATATCCTTTTGACACCAGATGGGACTGCCAAGGTCACAGACTTTGGGATTGCTGTAGCCTTTGCAGAGACAAGTCTG
ACCCAGACTAACTCGATGTTGGGCTCAGTTCATTACTTGTCCCCAGAGCAGGCGCGTGGTTCGAAGGCGACTGTGCAGAGTGA
TATCTATGCTATGGGGATTATTTTCTATGAGATGTTGACAGGCCATATCCCTTATGACGGGGATAGCGCGGTGACCATTGCCC
TCCAGCATTTCCAGAAACCCCTGCCGTCCGTTATTGCAGAAAATCCATCGTACCTCAGGCTTTAGAAAATGTTATTATCAAG
GCAACTGCTAAAAAGTTGACCAATCGCTACCGCTCGGTTTCAGAGATGTATGTGGACTTGTCTAGTAGCTTGTCCTACAATCG
TAGAAATGAAAGTAAGTTAATCTTTGATGAAACGAGCAAGGCAGATACCAAGACCTTGCCGAAGGTTTCTCAGAGTACCTTGA
CATCTATTCCTAAGGTTCAAGCGCAGACAGAACACAAATCAATCAAAAACCCAAGCCGGGCTGTGACAGAGGAAACTTACCAA
CCACAAGCACCGAAAAAACATAGATTTAAGATGCGTTACCTGATTTTGTTGGCCAGCCTTGTATTGGTGGCAGCTTCTCTTAT
TTGGATACTATCCAGAACTCCTGCAACCATTGCCATTCCAGATGTGGCAGGTCAGACAGTTGCAGAGGCCAAGGCAACGCTCA
AAAAAGCCAATTTTGAGATTGGTGAGGAGAAGACAGAGGCTAGTGAAAAGGTGGAAGAAGGGCGGATTATCCGTACAGATCCT
GGCGCTGGAACTGGTCGAAAAGAAGGAACGAAAATCAATCTGGTTGTCTCATCAGGCAAACAATCCTTCCAAATTAGTAATTA
TGTCGGCCGGAAATCTTCTGATGTTATCGCGGAATTAAAAGAGAAAAAAGTTCCAGATAATTTGATTAAAATTGAGGAAGAAC
AGTCGAATGAGAGTGAGGCTGGAACGGTCCTGAAGCAAAGTCTACCAGAAGGTACGACCTATGACTTGAGCAAGGCAACTCAA
ATTGTTTTGACAGTAGCTAAAAAAGCTACGACGATTCAATTAGGGAACTATATTGGACGGAACTCTACAGAAGTAATCTCAGA
ACTCAAGCAGAAGAAGGTTCCTGAGAATTTGATTAAGATAGAGGAAGAAGAGTCCAGCGAAAGCGAACCAGGAACGATTATGA
AACAAAGTCCAGGTGCCGGAACGACTTATGATGTGAGTAAACCTACTCAAATTGTCTTGACAGTAGCTAAAAAAGTTACAAGT
GTTGCCATGCCGAGTTACATTGGTTCCAGCTTGGAGTTTACTAAGAACAATTTGATTCAAATTGTTGGGATTAAGGAAGCTA
TATAGAAGTTGTAGAAGTGACGACAGCGCCTGCAGGTAGTGCAGAAGGCATGGTTGTTGAACAAAGTCCTAGAGCAGGTGAAA
AGGTAGACCTCAATAAGACTAGAGTCAAGACTTCAATCTACAAACCTAAAACAACTTCAGCTACTCCT

SeqID 294
ATGATCCAAATCGGCAAGATTTTTGCCGGACGCTATCGGATTGTCAAACAGATTGGTCGAGGAGGCATGGCGGATGTCTACCT
AGCCAAAGACTTAATCTTAGATGGGGAAGAAGTGGCAGTGAAGGTTCTGAGGACCAACTACCAGACGGACCCGATAGCTGTAC
CTCGTTTTCAGCGTGAAGCGAGAGCTATGGCAGATCTAGACCATCCTCATATCGTTCGGATAACAGATATTGGTGAGGAAGAC
GGTCAACAGTATCTTGCAATGGAGTATGTTGCTGGACTAGACCTCAAACGCTATATCAAGGAACATTATCCTCTTTCTAATGA
AGAAGCAGTCCGTATCATGGGACAAATTCTCTTGGCTATGCGCTTGGCCCATACTCGAGGAATTGTTCACAGGGACTTGAAAC
CTCAAAATATCCTTTTGACACCAGATGGGACGCGCCAAGGTCACAGACTTTGGGATTGCTGTAGCCTTTGCAGAGACAAGTCTC
ACCCAGACTAACTCGATGTTGGGCTCAGTTCATTACTTGTCACCAGAGCAGGCGCGTGGTTCGAAGGCGACTGTGCAGAGTGA
TATCTATGCCATGGGGATTATTTTCTATGAGATGTTGACAGGCCATATCCCTTATGACGGGGATAGCGCGGTGACCATTGCCC
TCCAGCATTTCCAGAAACCCCTGCCGTCCGTTATTGCAGAAAATCCATCTGTACCTCAGGCTTTAGAAAATGTTATTATCAGC
GCAACTGCTAAAAAGTTGACCAATCGCTATCGCTCGGTTTCAGAGATGTATGTAGACTTGTCTAGTAGCTTGTCCTACAATCC
TAGAAATGAAAGTAAGTTAATCTTTGATGAAACGAGCAAGGCAGATACCAAGACCTTGCCGAAGGTTTCTCAGAGTACCTTGA
CATCTATTCCTAAGGTTCAAGCGCAGACAGGACACAAATCAATCAAAAACCCAAGCCAGGCTGTGACAGAGGAAACTTACCAA
CCACAAGCACCGAAAAAACATAGATTTAAGATGCGTTACCTGATTTTGTTGGCCAGCCTTGTATTGGTGGCAGCTTCTCTTAT
TTGGATACTATCCAGAACTCCTGCAACCATTGCCATTCCAGATGTGGCAGGTCAGACAGTTGCAGAGGCCAAGGCAACGCTCA
AAAAAGCCAATTTTGAGATTGGTGAGGAGAAGACAGAGGCTAGTGAAAAGGTGGAAGAAGGGCGGATTATCCGTACAGATCCT
GGCGCTGGAACTGGTCGAAAAGAAGGAACGAAAATTAATCTGGTTGTCTCATCAGGCAAACAATCCTTCCAAATTAGTAATTA
TGTCGGCCGGAAATCTTCTGATGTTATCGCGGAATTAAAAGAGAAAAAAGTTCCAGATAATTTGATTAAAATTGAGGAAGAAC
AGTCGAATGAGAGTGAGGCTGGAACGGTCCTGAAGCAAAGTCTACCAGAAGGTACGACCTATGACTTGAGCAAGGCAACTCAA
ATTATTTTGACAGTAGCTAAAAAAGCTACGACGATTCAATTAGGGAACTATATTGGACGGAACTCTACAGAAGTAATCTCAGA
ACTCAAGCAGAAGAAGGTTCCTGAGAATTTGATTAAGATAGAGGAAGAAGAGTCCAGCGAAAGCGAACCAGGAACGATTATGA
AACAAAGTCCAGGTGCCGGAACGACTTATGATGTGAGTAAACCTACTCAAATTGTCTTGACAGTAGCTAAAAAAGTTACAAGT
GTTGCCATGCCGAGTTACATTGGTTCCAGCTTGGAGTTTACTAAGAACAATTTGATTCAAATTGTTGGGATTAAGGAAGCTA
TATAGAAGTTGTAGAAGTGACGACAGCGCCTGCAGGTAGTGCAGAAGGCATGGTTGTTGAACAAAGTCCTAGAGCAGGTGAAA
AGGTAGACCTAAATAAGACTAGAGTCAAGATTTCAATCTACAAACCTAAAACAACTTCAGCTACTCCT

SeqID 295
ATGATCCAAATCGGCAAGATTTTTGCCGGACGCTATCGGATTGTCAAACAGATTGGTCGAGGAGGCATGGCGGATGTCTACC
AGCCAAAGACTTAATCTTAGATGGGGAAGAAGTGGCAGTGAAGGTTCTGAGGACCAACTACCAGACGGACCCGATAGCTGTA
CTCGTTTTCAGCGTGAAGCGAGAGCTATGGCAGATCTAGACCATCCTCATATCGTTCGGATAACAGATATTGGTGAGGAAGA
GGTCAACAGTATCTTGCAATGGAGTATGTTGCTGGACTAGACCTCAAACGCTATATCAAGGAACATTATCCTCTTTCTAATG
AGAAGCAGTCCGTATCATGGGACAAATTCTCTTGGCTATGCGCTTGGCCCATACTCGAGGAATTGTTCACAGGGACTTGAAA
CTCAAAATATCCTTTTGACACCAGATGGGACTGCCAAGGTCACAGACTTTGGGATTGCTGTAGCCTTTGCAGAGACAAGTCT
ACCCAGACTAACTCGATGTTGGGCTCAGTTCATTACTTGTCACCAGAGCAGGCGCGTGGTTCGAAGGCGACTGTGCAGAGTG
TATCTATGCCATGGGGATTATTTTCTATGAGATGTTGACAGGCCATATCCCTTATGACGGGGATAGCGCGGTGACCATTGCC
TCCAGCATTTCCAGAACCCCCTGCCGTCCGTTATTGCAGAAAATTCATCTGTACCTCAGGCTTTAGAAAATGTTATTATCAA
GCAACTGCTAAAAAGTTGACCAATCGCTATCGCTCGGTTTCAGAGATGTATGTAGACTTGTCTAGTAGCTTGTCCTACAATC
TAGAAATGAAAGTAAGTTAATCTTTGATGAAACGAGCAAGGCAGATACCAAGACCTTGCCGAAGGTTTCTCAGAGTACCTTG
CATCTATTCCTAAGGTTCAAGCGCAGACAGAACACAAATCAATCAAAAACCCAAGCCAGGCTGTGACAGAGGAAACTTACCA

CCACAAGCACCGAAAAAACATAGATTTAAGATGCGTTACCTGATTTTGTTGGCCAGCCTTGTATTGGTGGCAGCTTCTCTTAT
TTGGATACTATCCAGAACTCCTGCAACCATTGCCATTCCAGATGTGGCAGGTCAGACAGTTGCAGAGGCCAAGGCAACGCTCA
AAAAAGCCAATTTTGAGATTGGTGAGGAGAAGACAGAGGCTAGTGAAAAGGTGGAAGAAGGGCGGATTATCCGTACAGATCCT
GGCGCTGGAACTGGTCGAAAAGAAGGAACGAAAATTAATCTGGTTGTCTCATCAGGCAAACAATCCTTCCAAATTAGTAATTA
TGTCGGCCGGAAATCTTCTGATGTTATCGCGGAATTAAAAGAGAAAAAAGTTCCAGATAATTTGATTAAAAATTGAGGAAGAAG
AGTCGAATGAAAGTGAGGCTGGAACGGTCCTGAAGCAAAGTCTACCAGAAGGTACGACCTATGACTTGAGCAAGGCAACTCAA
ATTGTTTTGACAGTAGCTAAAAAAGCTACGACGATTCAATTAGGGAACTATATTGGACGGAACTCTACAGAAGTAATCTCAGA
ACTCAAGCAGAAGAAGGTTCCTGAGAATTTGATTAAGATAGAGGAAGAAGAGTCCAGCGAAAGCGAACCAGGAACGATTATGA
AACAAAGTCCAGGTGCCGGAACGACTTATGTGAGTAAACCTACTCAAATTGTCTTGACAGTAGCTAAAAAAGTTACAAGT
GTTGCCATGCCGAGTTACATTGGTTCCAGCTTGGAGTTTACTAAGAACAATTTGATTCAAATTGTGGGATTAAGGAAGCTAA
TATAGAAGTTGTAGAAGTGACGACAGCGCCTGCAGGTAGTGTAGAAGGCATGGTTGTTGAACAAAGTCCTAGAGCAGGTGAAA
AGGTAGACCTAAATAAGACTAGAGTCAAGATTTCAATCTACAAACCTAAAACAACTTCAGCTACTCCTTAA

SeqID 296
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTATTGGAGTAGCTAGTGTAGCTGTTGCCAG
TCTTGTTATGGGAAGTGTGGTTCATGCGACAGAGAACGAGGGAAGTACCCAAGCAGCCACTTCTTCTAATATGGCAAAGACAG
AACATAGGAAAGCTGCTAAACAAGTCGTCGATGAATATATAGAAAAAATGTTGAGGGAGATTCAACTAGATAGAAGAAAACAT
ACCCAAAATGTCGCCTTAAACATAAAGTTGAGCGCAATTAAAACGAAGTATTTGCGTGAATTAAATGTTTTAGAAGAGAAGTC
GAAAGATGAGTTGCCGTCAGAAATAAAAGCAAAGTTAGACGCAGCTTTTGAGAAGTTTAAAAAAGATACATTGAAACCAGGAC
AAAAGGTAGCAGAAGCTAAGAAGAAGGTTGAAGAAGCTAAGAAAAAAGCCGAGGATCAAAAAGAAGAAGATCGTCGTAACTAC
CCAACCAATACTTACAAAACGCTTGAACTTGAAATTGCTGAGTTCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACTAGTAAA
AGAGGAAGCTAAAGAATCTCGAAACGAGGGCACAATTAAGCAAGCAAAAGAGAAAGTTGAGAGTAAAAAAGCTGAGGCTACAA
GGTTAGAAAAACATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCAGATGCTAAGTTGAAGGAAGCTAAT
GTAGCGACTTCAGATCAAGGTAAACCAAAGGGGCGGGCAAAACGAGGAGTTCCTGGAGAGCTAGCAACACCTGATAAAAAAGA
AAATGATGCGAAGTCTTCAGATTCTAGCGTAGGTGAAGAAACTCTTCCAAGCTCATCCCTGAAATCAGGAAAAAAGGTAGCAC
AAGCTGAGAAGAAGGTTGAAGAAGCTGAGAAAAAAGCCAAGGATCAAAAAGAAGAAGATCGCCGTAACTACCCAACCAATACT
TACAAAACGCTTGACCTTGAAATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACTAGTAAAAGAGGAAGCTAA
GGAACCTCGAGACGAGGAAAAAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAAGCTGAGGCTACAAGGTTAGAAAACA
TCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCAGCAGAAGAAGATAAAGTTAAAGAAAAACCAGCTGAA
CAACCACAACCAGCGCCGGCTACTCAACCAGAAAAACCAGCTCCAAAACCAGAGAAGCCAGCTGAACAACCAAAAGCAGAAA
AACAGATGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTGACTCAACAGCAACCGCCA
AAACTGAAAAACCAGCACAACCATCTACTCCAAAAACAGGCTGGAAACAAGAAAACGGTATGTGGTACTTCTACAATACTGAT
GGTTCAATGGCAACAGGATGGCTCCAAAACAACGGTTCATGGTACTATCTAAACGCTAATGGTGCTATGGCGACAGGATGGCT
CCAAAACAATGGTTCATGGTACTATCTAAACGCTAATGGTTCAATGGCAACAGGATGGCTCCAAAACAATGGTTCATGGTACT
ACCTAAACGCTAATGGTGCTATGGCGACAGGATGGCTCCAATACAATGGTTCATGGTACTACCTAAACAGCAATGGCGCTATG
GCGACAGGATGGCTCCAATACAATGGCTCCAAAACAACGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAACA
CGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAATACAACGGTTCATGGTATTACCTCAACC
CTAATGGTGATATGGCGACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGG
CAATGGTTCAAAGTATCAGATAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATGGCTATGG
AGTCAATGCCAATGGTGAATGGGTAAACTAA

SeqID 297
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTGTTGGAGTAGCTAGTGTAGTTGTTGCCAG
TCTTGTTATGGGAAGTGTGGTTCATGCGACAGAGAACGAGGGAGCTACCCAAGTACCCACTTCTTCTAATAGGGCAAATGAAA
GTCAGGCAGAACAAGGAGAACAACCTAAAAAACTCGATTCAGAACGAGATAAGGCAAGGAAAGAGGTCGAGGAATATGTAAA
AAAATAGTGGGTGAGAGCTATGCAAAATCAACTAAAAAGCGACATACAATTACTGTAGCTCTAGTTAACGAGTTGAACAACA
TAAGAACGAGTATTTGAATAAAATAGTTGAATCAACCTCAGAAAGCCAACTACAGATACTGATGATGGAGAGTCGATCAAAA
TAGATGAAGCTGTGTCTAAGTTTGAAAAGGACTCACCTTCTTCGTCAAGTTCAGACTCTTCCACTAAACCGGAAGXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXTTGCTGAGTCCGATGTGGAAGTTAAAAAAGCGGAGCTTGAACTAGTAAAA
AGGAAGCTAAGGAACCTCGAAACGAGGAAAAAGTTAAGCAAGCAAAAGCGGAAGTTGAGAGTAAAAAAGCTGAGGCTACAAGG
TTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCAGCAGAAGAAGATAAAGTTAAAGAAA
ACCAGCTGAACAACCACAACCAGCGCCGGCTCCAAAAGCAGAAAAAACCAGCTCCAGCTCCAAAACCAGAGAATCCAGCTGAA
AACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTGAC
CAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAAACAGGCTGGAAACAAGAAAACGGTATGTGGT
CTTCTACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAAAACAATGGCTCATGGTACTACCTCAACAGCAATGGCGCT
TGGCGACAGGATGGCTCCAAAACAATGGTTCATGGTACTATCTAAACGCTAATGGTTCAATGGCAACAGGATGGCTCCAAAA
AATGGTTCATGGTACTACCTAAACGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTACCTAA
CGCTAATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACA
GTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAGTATCAGA
AAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATGGCTATGGAGTCAATGCCAATGGTGAAT
GGTAAAC

SeqID 298
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTGAAATTGCTGAGTCCGATGTGGAAGTTAAAAAAGCGGAG
TTGACTAGTAAAGAGGAAGCTAAGAAGCCTTTAAACGAGGGCACAATTAGGCAAGCAAAAGCGGAAGTTGAGAGTAAAAAAG
TGAGGCTACAAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAGAGCAGCAGAAGAAGAT
AAGTTAAAGAAAAACCAGCTGAACAACCACAACCAGCGCCGGCCCCTCAACCAGAAAAACCAACTGAAGAGCCTGAGAATCC
GTTCCAGCTCCAAAACCAGAGAAGCCAGCTGAACAACCAAAAACCAGAGAAGCCAGCTGAACAACCAAAACCAGAGAAGCCAG

```
TGAACAACCAAAAGCAGAAAAAACAGATGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCT
TGACTCAACAGCAACCGCAAAAACCAGAGCAACCAGCTCCTGCACCAAAAATAGGCTGGAAACAAGAAAACGGTATGTGGTAC
TTCTACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTATTACCTAAACGCTAACGGTTCAAT
GGCAACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAG
TATCAGATAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATGGCTATGGAGTCAATGCCAAT
GGTGAATGGGTAAAC
```

SeqID 299
```
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTGTTGGAGTAGCTAGTGTAGCTGTTGCCAG
TCTTGTTATGGGAAGTGTGGTTCATGCGACAGAGAACGAGAGAACTACCCAAGTACCCACTTCTTCTAATAGGGGAAAGCCAG
AACGTAGGAAAGCTGCTGAACAATTCGATGAATATATAAACAAAATGATCCAATTAGATAAAAGAAAACATACCCAAAATTTA
GCCTTCAACATACAGTTGAGCAGAATTAAAACGGAGTATTTGAATGGATTAAAAGAGAAGTCGGAAGCTGAGTTGCCGTCAAA
AATAAAAGCAGAGTTAGACGCAGCTTTTAAGCAGTTTAAAAAAGATACATTACCAACAGAACCAGAAAAAAAGTAGCAGAAG
CTGAGAAGAAGGTTGAAGAAGCTGAGAAGAAGGTAGCAGAAGCTAAGAAAAAAGCCAAGGCTCAAAAAGAAGAAGATCACCGT
AACTACCCAACCATTACTTACAAAACGCTTGACCTTGAAATTGCTGAGTTCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACT
AGTAAAAAAGGAAGCTGACGAATCTCGAAACGAGGGCACAATTAACCAAGCAAAAGCGAAAGTTGAGAGTGAAAAAGCTGAGC
CTACAAGGTTAAAAAAAATCAAGACAGATCGTGAAAAAGCAGAAGAAGAAGCTAAACGAAGAGCAGATGCTAAAGAGCAA
GATGAATCAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGG
AGCTTGAACTAGTAAAAGAGGAAGCTAAGGAATCTCGAAACGAGGAAAAAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAA
AAAGCTGAGGCTACAAGGTTAGAAAAAAATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCAGCAGAAGA
AGATAAAGTTAAAGAAAAACCAGCTGAACAACCACAACCAGCGCCGGCTCCAAAACCAGAGAATCCAGCTGAAGAGCCTGAGA
ATCCAGTTCCAGCTCCAAAACCAGAGAATCCAGCTGAACAACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTGAAGAAGAC
TATGCTCGTAGATCAGAAGAAGAATATAATCGCTTGACTCAACAGCAACCGCCAAAACCAGAGCAACCAGCTCCTGCACCAAA
AATAGGCTGGAAACAAGAAAACGGTATGTGGTACTTCTACAATACTGATGGTTCAATGGCGACCGGATGGCTCCAAAACAATG
GCTCATGGTACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCCAAAACAATGGTTCATGGTACTATCTAAACGCT
AATGGTTCAATGGCAACAGGATGGCTCCAAAACAATGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATG
GCTCCAATACAATGGCTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAATACAATGGCTCATGGT
ACTACCTAAACGCTAATGGTGATATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTATCTAAACGCTAATGGTGA
ATGGCGACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCA
AGTATCAGATAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGACGGCTATGGAGTCAATGCCA
ATGGTGAATGGGTAAAC
```

SeqID 300
```
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTATTGGAGTAGCTAGTGTAGCTGTTGCTAC
CTTGTTCTTAGGAGGAGTAGTCCATGCAGAAGGGGTTAGAAGTGAGAATACCCCCAAGGTTACATCTAGTGGGGATGAAGTCC
ATGAATATATAAAAAAAATGTTGAGTGAGATCCAATTAGATAAAAGAAAACATACCCACAATTTCGCCTTAAACCTAAAGTTC
AGCAGAATTAAAACGGAGTATTTGTATAAATTAAAAGTTAATGTTTTAGAAGAAAAGTCAAAAGCTGAGTTGACGTCAAAAAC
AAAAAAAGAGGTAGACGCAGCTTTTGAGAAGTTTAAAAAAAGATACATTGAAACTAGGAGAAAAGGTAGCAGAAGCACAGAAGA
AGGTTGAAGAAGCTAAGAAAAAAGCCAAGGATCAAAAAGAAGAAGATCACCGTAACTACCCAACCAATACTTACAAAACGCTT
GAACTTGAAATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACTATTGAAAGAGGAAGCTAAAACTCGAAACGA
GGACACAATTAACCAAGCAAAAGCGAAAGTTAAGAGTGAACAAGCTGAGGCTACAAGGTTAAAAAAAATCAAGACAGATCGTC
AACAAGCTGAGGCTACAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXATTGCTGAGTCCGATGTGAAAGTTA
AGAAGCGGAGCTTGAACTAGTAAAAGAGGAAGCTAAGGAACCTCGAGACGAGGAAAAAATTAAGCAAGCAAAAGCGGAAGTTC
AGAGTAAACAAGCTGAGGCTACAAGGTTAGAAAAAAATCAAGACAGATCGTGAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCA
GCAGAAGAAGATAAAGTTAAAGAAAAACCAGCTGAACAACCACAACCAGCGCCGGCTCCTCAACCAGAAAAACCAACTCCAA
ACCAGAAAAACCAGCTCCAGCTCCAAAACCAGAGAATCCAGCTGAACAACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTC
AAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTGACTCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAA
CCATCTACTCCAAAAACAGGCTGGAAACAAGAAAACGGTATGTGGTACTTCTACAATACTGATGGTTCAATGGCGACAGGATC
GCTCCAATACAATGGCTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAACAATGGTTCATGG
ACTACCTAAACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAACAATGGTTCATGGTACTACCTAAACGCTAATGGTGA
ATGGCGACAGGATGGCTCCAAAACAATGGCTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACAGGTTGGGTGAAAG
TGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAGCATCAGATAAATGGTACTATC
TCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATGGCTATGGAGTCAATGCCAATGGTGAATGGGTAAAC
```

SeqID 301
```
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTATTGGAGTAGCTAGTGTAGCTGTTGCCAC
TCTTGTTATGGGAAGTGTGGTTCATGCGACAGAGAAGGAGGTAACTACCCAAGTACCCACTTATTCTAATATGGCAAAGACAC
AACATAGGAAAGCTGCTAAACAAGTCGTCGATGAATATATAGAAAAAATGTTGAGGGAGATTCAATTAGATAGAAGAAAACA
ACCCAAAATTTCGCCTTCAACATGAAGTTGAGCGCAATTAAAACGGAGTATTTGTATGGATTAAAAGAGAAGTCGGAAGCTGA
GTTGCCGTCATCGGAAGCTGAGTTGCCGTCAGAAGTAAAAGCAAAGTTAGACGCAGCTTTTGAGCAGTTTAAAAAAGATACAT
TGAAACTAGGAGAAAAGGTAGCAGAAGCTGAGAAGAAGGTTGCAGAAGCTGAGAAAAAGCCAAGGCTCAAAAAGAAGAAGA
CGCCGTAACTACCCAACCATTACTTACAAAACGCTTGAAATTGCTGAGTCCGATGTGAAGTTAAAAAAGCGGAGC
TGAACTATTGAAAGAGGAAGCTAAAACTCGAAACAAGGACACAATTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAAGCTC
AGGCTACAAAGTTAGAAGAAATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACTAGTAAAAGAGGAAGCTAAGGAATC
CGAAACGAGGAAAAAGTTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAAAAAGCTGAGGCTACAAGGTTAGAAAAAATCAAGA
AGATCGTAAAAAAGCAGAAGAAGCTAAACGAAGAGCAGCAGAAGAAGATAAAGTTAAAGAAAAACCAGCTGAACAACCACAA
CAGCGCCGGCTCCTCAACCAGAAAAACCAACTGAAGAGCCTGAGAATCCAGCTCCAGCTCCAAAACCTGAGAAACCAGCTGA
```

CAACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAATCGCTTGAC
TCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAAACAGGCTGGAAACAAGAAAACGGTATGTGGT
ACTTCTACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAAAACAATGGCTCATGGTACTACCTCAACAGCAATGGCGCT
ATGGCGACAGGATGGCTCCAATACAATGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAA
CAATGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAATACAATGGTTCATGGTACTACCTCA
ACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAACAATGGCTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACA
GGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAGTATCAGA
TAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATGGCTATGGAGTCAATGCCAATGGTGAAT
GGGTAACC

SeqID 302
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTATTGGAGTAGCTAGTGTAGCTGTTGCTAG
CTTGTTCTTAGGAGGAGTAGTCCATGCAGAAGGGGTTAGAAGTGAGAATACCCCCAAGGTTACATCTAGTGGGGATGAAGTCG
ATGAATATATAAAAAAAATGTTGAGTGAGATCCAATTAGATAAAAGAAAACATACCCACAATTTCGCCTTAAACCTAAAGTTG
AGCAGAATTAAAACGGAGTATTTGTATAAAATTAAAAGTTAATGTTTTAGAAGAAAAAGTCAAAAGCTGAGTTGACGTCAAAAC
AAAAAAAGAGGTAGACGCAGCTTTTGAGAAGTTTAAAAAAAGATACATTGAAACTAGGAGAAAAGGTAGCGAAGCTCAGAAGA
AGGTTGAAGAAGCTAAGAAAAAAGCCAAGGATCAAAAGAAGAAGATCACCGTAACTACCCAACCAATACTTACAAAACGCTT
GAACTTGAAATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGGAGCTTGAACTATTGAAAGAGGAAGCTAAAACTCGAAACGA
GGACACAATTAACCAAGCAAAAGCGAAAGTTAAGAGTGAACAAGCTGAGGCTACAAGGTTAAAAAAAATCAAGACAGATCGTG
AACAAGCTGAGGCTACAAGGTTAGAAAACATCAAGACAGATCGTGAAAAAGCAGAAGAAGCTAAACGAAAAGCAGAAXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTTGAAATTGCTGAGTCCGATGTGAAAGTTAAAGAAGCGGAGCTTGAAC
TAGTAAAGAGGAAGCTAAGGAATCTCGAAACGAGGAAAAAGTTAAGCAAGCAAAAGCGAAAGTTGAGAGTAAACAAGCTGAC
GCTACAAGGTTAGAAAAAATCAAGACAGATCGTAAAAAAGCAGAAGAAGAAGCTAAACGAAAAGCAGCAGAAGAAGATAAAGT
TAAAGAAAAACCAGCTGAACAACCACAACCAGCGCCGGCTCCTCAACCAGAAAAACCAGCTCCAGCTCCAAAACCAGAGAATC
CAGCTGAACAACCAAAAGCAGAAAAACCAGCTGATCAACAAGCTGAAGAAGACTATGCTCGTAGATCAGAAGAAGAATATAAT
CGCTTGACTCAACAGCAACCGCCAAAAACTGAAAAACCAGCACAACCATCTACTCCAAAAACAGGCTGGAAACAAGAAAACGG
TATGTGGTACTTCTACAATACTGATGGTTCAATGGCGACAGGATGGCTCCAATACAATGGCTCATGGTACTACCTAAACGCTA
ATGGTGATATGGCGACAGGATGGCTCCAAAACAATGGTTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACAGGATGC
CTCCAAAACAATGGCTCATGGTACTACCTAAACGCTAATGGTGATATGGCGACAGGATGGCTCCAAAACAATGGCTCATGGTA
CTACCTAAACGCTAATGGTGATATGGCGACAGGTTGGGTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTA
TGAAAGCAAGCCAATGGTTCAAAGCATCAGATAAATGGTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTA
GATGGCTATGGAGTCAATGCCAATGGTGAATGGGTAAAC

SeqID 303
ATGTTTGCATCAAAAAGCGAAAGAAAAGTACATTATTCAATTCGTAAATTTAGTATTGGAGTAGCTAGTGTAGCTGTTGCCAC
TCTTTTTATGGGAAGTGTGGTTCATGCGACAGAGAAGGAGGTAACTACCCAAGTAGCCACTTCTTCTAATAAGGCAAATAAAA
GTCAGACAGAACATATGAAAGCTGCTAAACAAGTCGATGAATATATAAAAAAAAAAGCTCCAATTAGATAGAAGAAAAACATACC
CAAAATGTCGGCTTACTCACAAAGTTGGGCGTAATTAAAACGGAGTATTTGCATGGATTAAGTGTTTCAAAAAAGAAGTCGGA
AGCTGAGTTGCCGTCAGAAATAAAAGCAAAGTTAGACGCAGCTTTTGAGCAGTTTAAAAAAAGATACATTACCAACAGAACCAC
GAAAAAAGGTAGCAGAAGCTGAGAAGAAGGTTGAAGAAGCTAAGAAAAAAGCCGAGGATCAAAAGAAAAAGATCTCCGTAAC
TACCCAACCAATACTTACAAAACGCTTGAACTTGACATTGCTGAGTCCGATGTGGAAGTTAAAAAAGCGGAGCTTGAACTAGT
AAAAGGGAGCTACAGGAATCTCGAGACGAGAAAAAAATTAATCAAGCAAAGCGAAAGTTGAGAATAAAAAAGCTGATGCTAC
AAGGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGTCCAAAACGGTATGTGGTACTTCTACAATACTGATGG
TCAATGGCAACAGGCTGGCTCCAAAACAATGGCTCATGGTACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCC
AAACAATGGCTCATGGTACTACCTCAACAGCAATGGCGCTATGGCGACAGGATGGCTCCAATACAATGGTTCATGGTACTAC
TCAACGCTAATGGTGATATGGCGACAGGATGGTTCCAATACAATGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGC
ACAGGATGGTTCCAATACAATGGTTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGTTCCAATACAATG
TTCATGGTACTACCTCAACGCTAATGGTGATATGGCGACAGGATGGCTCCAATACAATGGTTCATGGTACTACCTAAACAGC
ATGGTGCTATGGTAACAGGATGGCTCCAAAACAATGGCTCATGGTACTACCTAAACGCTAACGGTTCAATGGCAACAGATTG
GTGAAAGATGGAGATACCTGGTACTATCTTGAAGCATCAGGTGCTATGAAAGCAAGCCAATGGTTCAAAGTATCAGATAAAT
GTACTATGTCAATGGCTCAGGTGCCCTTGCAGTCAACACAACTGTAGATAGCTATAGAGTCAATGCCAATGGTGAATGGGTA
AC

SeqID 304
MKKKILASLLLSTVMVSQVAVLTTAHAETTDDKIAAQDNKISNLTAQQQEAQKQVDQIQEQVSAIQAEQSNLQAENDRLQAE
KKLEGEITELSKNIVSRNQSLEKQARSAQTNGAVTSYINTIVNSKSITEAISRVAAMSEIVSANNKMLEQQKADKKAISEKQ
ANNDAINTVIANQQKLADDAQALTTKQAELKAAELSLAAEKATAEGEKASLLEQKAAAEAEARAAAVAEAAYKEKRASQQQS
LASANTNLTAQVQAVSESAAAPVRAKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAQWATSAAAAGFRTGSTPQVG
IACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTTTPEGFVTYIYAD

SeqID 305
MKKKILASLLLSTVMVSQVAVLTTAHAETTDDKIAAQDNKISNLTAQQQEAQKQVDQIQEQVSAIQAEQSNLQAENDRLQAE
KKLEGEITELSKNIVSRNQSLEKQARSAQTNGAVTSYINTIVNSKSITEAISRVAAMSEIVSANNKMLEQQKADKKAISEKQ
ANNDAINTVIANQQKLADDAQALTTKQAELKAAELSLAAEKATAEGEKASLLEQKAAAEAEARAAAVAEAAYKEKRAGQQQS
LASANTNLTAQVQAVSESAAAPVRAKVRPTYSTNASSYPIGECTWGVKTLAPWAGDYWGNGAQWATSAAAAGFRTGSTPQVG
IACWNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGWFNPTTTSEGFVTYIYAD

SeqID 306

MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEEL
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAARIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSI
TQTNSMLGSVHYLSPEQAGGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQKPLPSVIAENPSVPQALENVIIF
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTEHKSIKNPSQAVTEETYC
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDF
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATC
IVLTVAKKVTSVAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSAEGMVVEQSPRAGEKVDLNKTRVKISIYKPF
TTSATP

SeqID 307
MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEEL
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAVRIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSI
TQTNSMLGSVHYLSPEQARGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQKPLPSVIAENPSVPQALENVIIF
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTEHKSIKNPSQAVTEETYC
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDF
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATC
IVLTVAKKATTIQLGNYIGRNSTEVISELKQKKVPENLIKIEEEESSESEPGTIMKQSPGAGTTYDVSKPTQIVLTVAKKVTS
VAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSVEGMVVEQSPRAGEKVDLNKTRVKISIYKPKTTSATP

SeqID 308
MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEEI
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAVRIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSI
TQTNSMLGSVHYLSPEQARGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQKPLPSVIAENPSVPQALENVIII
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTEHKSIKNPSRAVTEETYC
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDF
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATC
IVLTVAKKATTIQLGNYIGRNSTEVISELKQKKVPENLIKIEEEESSESEPGTIMKQSPGAGTTYDVSKPTQIVLTVAKKVTS
VAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSAEGMVVEQSPRAGEKVDLNKTRVKTSIYKPKTTSATP

SeqID 309
MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEEI
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAVRIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSI
TQTNSMLGSVHYLSPEQARGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQKPLPSVIAENPSVPQALENVIII
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTGHKSIKNPSQAVTEETYC
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDF
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATC
IILTVAKKATTIQLGNYIGRNSTEVISELKQKKVPENLIKIEEEESSESEPGTIMKQSPGAGTTYDVSKPTQIVLTVAKKVTS
VAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSAEGMVVEQSPRAGEKVDLNKTRVKISIYKPKTTSATP

SeqID 310
MIQIGKIFAGRYRIVKQIGRGGMADVYLAKDLILDGEEVAVKVLRTNYQTDPIAVARFQREARAMADLDHPHIVRITDIGEEI
GQQYLAMEYVAGLDLKRYIKEHYPLSNEEAVRIMGQILLAMRLAHTRGIVHRDLKPQNILLTPDGTAKVTDFGIAVAFAETSI
TQTNSMLGSVHYLSPEQARGSKATVQSDIYAMGIIFYEMLTGHIPYDGDSAVTIALQHFQNPLPSVIAENSSVPQALENVIII
ATAKKLTNRYRSVSEMYVDLSSSLSYNRRNESKLIFDETSKADTKTLPKVSQSTLTSIPKVQAQTEHKSIKNPSQAVTEETYC
PQAPKKHRFKMRYLILLASLVLVAASLIWILSRTPATIAIPDVAGQTVAEAKATLKKANFEIGEEKTEASEKVEEGRIIRTDI
GAGTGRKEGTKINLVVSSGKQSFQISNYVGRKSSDVIAELKEKKVPDNLIKIEEEESNESEAGTVLKQSLPEGTTYDLSKATC
IVLTVAKKATTIQLGNYIGRNSTEVISELKQKKVPENLIKIEEEESSESEPGTIMKQSPGAGTTYDVSKPTQIVLTVAKKVTS
VAMPSYIGSSLEFTKNNLIQIVGIKEANIEVVEVTTAPAGSVEGMVVEQSPRAGEKVDLNKTRVKISIYKPKTTSATP

SeqID 311
MFASKSERKVHYSIRKFSIGVASVAVASLVMGSVVHATENEGSTQAATSSNMAKTEHRKAAKQVVDEYIEKMLREIQLDRRKI
TQNVALNIKLSAIKTKYLRELNVLEEKSKDELPSEIKAKLDAAFEKFKDTLKPGEKVAEAKKKVEEAKKKAEDQKEEDRRN
PTNTYKTLELEIAEFDVKVKEAELELVKEEAKESRNEGTIKQAKEKVESKKAEATRLENIKTDRKKAEEEAKRKADAKLKEAI
VATSDQGKPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSSSLKSGKKVAEAEKKVEEAEKKAKDQKEEDRRNYPTNT
YKTLDLEIAESDVKVKEAELELVKEEAKEPRDEEKIKQAKAKVESKKAEATRLENIKTDRKKAEEEAKRKAAEEDKVKEKPA:
QPQPAPATQPEKPAPKPEKPAEQPKAEKTDDQQAEEDYARRSEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMWYFYNTI
GSMATGWLQNNGSWYYLNANGAMATGWLQNNGSWYYLNANGSMATGWLQNNGSWYYLNANGAMATGWLQYNGSWYYLNSNGAI
ATGWLQYNGSWYYLNANGDMATGWLQNNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMKA:
QWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 312
MFASKSERKVHYSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQGEQPKKLDSERDKARKEVEEYV.
KIVGESYAKSTKKRHTITVALVNELNNIKNEYLNKIVESTSESQLQILMMESRSKVDEAVSKFEKDSPSSSSSDSSTKPEXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAESDVEVKKAELELVKEEAKEPRNEEKVKQAKAEV.
SKKAEATRLEKIKTDRKKAEEEAKRKAAEEDKVKEKPAEQPQPAPAPKAEKPAPAPKPENPAEQPKAEKPADQQAEEDYARR:
EEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATGWLQNNGSWYYLNANG:
MATGWLQNNGSWYYLNANGSMATGWLQYNGSWYYLNANGSMATGWLQYNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMK:
SQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 313
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLKLLSPMWKLKKRSLTSKEEAKKPLNEGTIRQ
AKAEVESKKAEATRLEKIKTDRKKAEEEAKRRAAEEDKVKEKPAEQPQPAPAPQPEKPTEEPENPVPAPKPEKPAEQPKPEKP
AEQPKPEKPAEQPKAEKTDDQQAEEDYARRSEEEYNRLTQQQPQKPEQPAPAPKIGWKQENGMWYFYNTDGSMATGWLQYNGS
WYYLNANGSMATGWVKDGDTWYYLEASGAMKASQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 314
MFASKSERKVHYSIRKFSVGVASVAVASLVMGSVVHATENERTTQVPTSSNRGKPERRKAAEQFDEYINKMIQLDKRKHTQNL
AFNIQLSRIKTEYLNGLKEKSEAELPSKIKAELDAAFKQFKKDTLPTEPEKKVAEAEKKVEEAEKKVAEAKKKAKAQKEEDHR
NYPTITYKTLDLEIAEFDVKVKEAELELVKKEADESRNEGTINQAKAKVESEKAEATRLKKIKTDREKAEEEEAKRRADAKEQ
DESXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXIAESDVKVKEAELELVKEEAKESRNEEKIK
QAKAKVESKKAEATRLEKIKTDRKKAEEEAKRRAAEEDKVKEKPAEQPQPAPAPKPENPAEEPENPVPAPKPENPAEQPKAEK
PADQQAEEDYARRSEEEYNRLTQQQPPKPEQPAPAPKIGWKQENGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATGWLQN
NGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWVKDGDT
WYYLEASGAMKASQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 315
MFASKSERKVHYSIRKFSIGVASVAVASLFLGGVVHAEGVRSENTPKVTSSGDEVDEYIKKMLSEIQLDKRKHTHNFALNLKI
SRIKTEYLYKLKVNVLEEKSKAELTSKTKKEVDAAFEKFKKDTLKLGEKVAEAQKKVEEAKKKAKDQKEEDHRNYPTNTYKTI
ELEIAESDVKVKEAELELLKEEAKTRNEDTINQAKAKVKSEQAEATRLKKIKTDREQAEATXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXIAESDVKVKEAELELVKEEAKEPRDEEKIKQAKAEVESKQAEATRLEKIKTDRKK
AEEEAKRKAAEEDKVKEKPAEQPQPAPAPQPEKPTPKPEKPAPAPKPENPAEQPKAEKPADQQAEEDYARRSEEEYNRLTQQQ
PPKTEKPAQPSTPKTGWKQENGMWYFYNTDGSMATGWLQYNGSWYYLNANGDMATGWLQNNGSWYYLNANGDMATGWLQNNGS
WYYLNANGDMATGWLQNNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMKASQWFKASDKWYYVNGSGALAVNTTVDGYGVN
ANGEWVN

SeqID 316
MFASKSERKVHYSIRKFSIGVASVAVASLVMGSVVHATEKEVTTQVPTYSNMAKTEHRKAAKQVVDEYIEKMLREIQLDRRKK
TQNFAFNMKLSAIKTEYLYGLKEKSEAELPSSEAELPSEVKAKLDAAFEQFKKDTLKLGEKVAEAEKKVAEAEKKAKAQKEEI
RRNYPTITYKTLDLEIAESDVEVKKAELELLKEEAKTRNKDTIKQAKAKVESKKAEATKLEEIKTDRKKAEEEAXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXIAESDVKVKEAELELVKEEAKESRNEEKVKQAKAKVESKKAE
ATRLEKIKTDRKKAEEEAKRRAAEEDKVKEKPAEQPQPAPAPQPEKPTEEPENPAPAPKPEKPAEQPKAEKPADQQAEEDYARF
SEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATGWLQYNGSWYYLNANG
DMATGWLQNNGSWYYLNANGDMATGWLQYNGSWYYLNANGDMATGWLQNNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMK
ASQWFKVSDKWYYVNGSGALAVNTTVDGYGVNANGEWVN

SeqID 317
MFASKSERKVHYSIRKFSIGVASVAVASLFLGGVVHAEGVRSENTPKVTSSGDEVDEYIKKMLSEIQLDKRKHTHNFALNLKI
SRIKTEYLYKLKVNVLEEKSKAELTSKTKKEVDAAFEKFKKDTLKLGEKVAEAQKKVEEAKKKAKDQKEEDHRNYPTNTYKTI
ELEIAESDVKVKEAELELLKEEAKTRNEDTINQAKAKVKSEQAEATRLKKIKTDREQAEATRLENIKTDREKAEEAKRKAEXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXLEIAESDVKVKEAELELVKEEAKESRNEEKVKQAF
AKVESKQAEATRLEKIKTDRKKAEEEAKRKAAEEDKVKEKPAEQPQPAPAPQPEKPAPAPKPENPAEQPKAEKPADQQAEED\
ARRSEEEYNRLTQQQPPKTEKPAQPSTPKTGWKQENGMWYFYNTDGSMATGWLQYNGSWYYLNANGDMATGWLQNNGSWYYLN
ANGDMATGWLQNNGSWYYLNANGDMATGWLQNNGSWYYLNANGDMATGWVKDGDTWYYLEASGAMKASQWFKASDKWYYVNGS
GALAVNTTVDGYGVNANGEWVN

SeqID 318
MFASKSERKVHYSIRKFSIGVASVAVASLFMGSVVHATEKEVTTQVATSSNKANKSQTEHMKAAKQVDEYIKKKLQLDRRKH\
QNVGLLTKLGVIKTEYLHGLSVSKKKSEAELPSEIKAKLDAAFEQFKKDTLPTEPGKKVAEAEKKVEEAKKKAEDQKEKDLRN
YPTNTYKTLELDIAESDVEVKKAELELVKGSYRNLETRKKLIKQSEKLRIKKLMLQXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXVQNGMWYFYNTDGSMATGWLQNNGSWYYLNSNGAMATGWLQNNGSWYYLNSNGAMATGWI
QYNGSWYYLNANGDMATGWFQYNGSWYYLNANGDMATGWFQYNGSWYYLNANGDMATGWFQYNGSWYYLNANGDMATGWLQYN
GSWYYLNSNGAMVTGWLQNNGSWYYLNANGSMATDWVKDGDTWYYLEASGAMKASQWFKVSDKWYYVNGSGALAVNTTVDSYI
VNANGEWVN

Note: "X" represents undefined/missing nucleotides or amino acides due t
unavailable sequencing information.

(54) Title: S. PNEUMONIAE ANTIGENS

(57) Abstract: The present invention discloses isolated nucleic acid molecules encoding a hyperimmune serum reactive antigen or a fragment thereof as well as hyperimmune serum reactive antigens or fragments thereof from S. pneumoniae, methods for isolating such antigens and specific uses thereof.

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7	C07K14/315	C07K16/12	C07K19/00	C07K16/46	C12N5/12
	C12N15/02	A61K39/09	C12Q1/68	G01N33/569	A61P31/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7   C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, EMBL, EMBASE, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02/083855 A (CHAKRAVARTI DEB NARAYAN ; RUSSELL DAVID PARRISH (US); WOOTERS JOSEPH L) 24 October 2002 (2002-10-24) page 85, line 15 - page 108, line 4 claims 1-11,25-28,59-71,96,105 see in particular SEQ ID Nos. 209 and 424.	1,2, 5-11, 14-37
X	WO 02/077021 A (CHIRON SPA ; MASIGNANI VEGA (IT); FRASER CLAIRE (US); TETTELIN HERVE () 3 October 2002 (2002-10-03) page 8, line 30 - page 34, line 40 see in particular SEQ ID Nos. 4715 and 4716. claims 1-16	1,2, 5-11, 14-37

-/--

X	Further documents are listed in the continuation of box C.		X	Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the International filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the International filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the International search	Date of mailing of the international search report
25 August 2004	25. 10. 2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Ulbrecht, M

7

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HOSKINS J ET AL: "Genome of the Bacterium Streptococcus pneumoniae Strain R6" JOURNAL OF BACTERIOLOGY, WASHINGTON, DC, US, vol. 183, no. 19, October 2001 (2001-10), pages 5709-5717, XP002231307 ISSN: 0021-9193 cited in the application the whole document	1,2, 5-11, 14-37
X	-& DATABASE EMBL EBI; 7 September 2001 (2001-09-07), HOSKINS ET AL.: "Streptococcus pneumoniae R6 section 1" XP002293255 retrieved from HTTP://WWW.EBI.AC.UK Database accession no. AE008385 cited in the application the whole document	1,2, 5-11, 14-37
X	TETTELIN H ET AL: "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 293, no. 5529, 2001, pages 498-506, XP002218261 ISSN: 0036-8075 cited in the application the whole document	1,2, 5-11, 14-37
X	-& DATABASE EMBL EBI; 31 July 2001 (2001-07-31), TETTELIN ET AL.: "Streptococcus pneumoniae TIGR4 section 1" XP002293256 retrieved from HTTP://WWW.EBI.AC.UK Database accession no. AE007318 cited in the application the whole document	1,2, 5-11, 14-37
A	NABORS G S ET AL: "Immunization of healthy adults with a single recombinant pneumococcal surface protein A (PspA) variant stimulates broadly cross-reactive antibodies to heterologous PspA molecules" VACCINE, BUTTERWORTH SCIENTIFIC. GUILDFORD, GB, vol. 18, no. 17, March 2000 (2000-03), pages 1743-1754, XP004190054 ISSN: 0264-410X the whole document	1,2, 5-11, 14-37

-/--

7

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BETHE GESINA ET AL: "The cell wall-associated serine protease PrtA: A highly conserved virulence factor of Streptococcus pneumoniae" FEMS MICROBIOLOGY LETTERS, vol. 205, no. 1, 27 November 2001 (2001-11-27), pages 99-104, XP002293235 ISSN: 0378-1097 the whole document	1,2, 5-11, 14-37
A	WIZEMANN T M ET AL: "Use of a whole genome approach to identify vaccine molecules affording protection against Streptococcus pneumoniae infection" INFECTION AND IMMUNITY, AMERICAN SOCIETY FOR MICROBIOLOGY. WASHINGTON, US, vol. 69, no. 3, March 2001 (2001-03), pages 1593-1598, XP002263818 ISSN: 0019-9567 the whole document	1,2, 5-11, 14-37
A	OVERWEG K ET AL: "The putative proteinase maturation protein A of Streptococcus pneumoniae is a conserved surface protein with potential to elicit protective immune responses" INFECTION AND IMMUNITY, AMERICAN SOCIETY FOR MICROBIOLOGY. WASHINGTON, US, vol. 68, no. 7, July 2000 (2000-07), pages 4180-4188, XP002158137 ISSN: 0019-9567 the whole document	1,2, 5-11, 14-37
A	WO 02/059148 A (CISTEM BIOTECHNOLOGIES GMBH ; AHSEN UWE (AT); ETZ HILDEGARD (AT); HAFN) 1 August 2002 (2002-08-01) the whole document	1,2, 5-11, 14-37

7

INTERNATIONAL SEARCH REPORT

Box II   Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
   because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
   because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. ☐ Claims Nos.:
   because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box III   Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

    see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

    see annex

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (January 2004)

.FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

This International Searching Authority found multiple (groups of)
inventions in this international application, as follows:

Invention 1: claims 1,2,5-11,14-37 (all partially)

> An isolated nucleic acid molecule encoding a hyperimmune
> serum-reactive antigen or fragment thereof as defined in
> claim 1, but referring only to SEQ ID No. 1; a vector
> comprising said nucleic acid molecule; a host cell
> comprising said vector; a hyperimmune serum-reactive antigen
> comprising the amino acid sequence of SEQ ID No. 145; a
> fragment of the said hyperimmune serum-reactive antigen as
> defined in claim 14; a method of producing the said S.
> pneumoniae hyperimmune serum-reactive antigen or fragment
> thereof; a process for producing a cell which expresses the
> said S. pneumoniae hyperimmune serum-reactive antigen or
> fragment thereof; the use of the said nucleic acid molecule
> or the said hyperimmune serum-reactive antigen or fragment
> thereof for the manufacture of a pharmaceutical preparation;
> an antibody, or at least an effective part thereof which
> binds at least to a selective part of the said hyperimmune
> serum-reactive antigen or fragment thereof; a hybridoma cell
> line which produces the said antibody; a method of producing
> the said antibody; the use of said antibody for the
> preparation of a medicament for treating or preventing S.
> pneumoniae infections; an antagonist which binds to the said
> hyperimmune serum-reactive antigen or fragment thereof; a
> method of identifying an antagonist capable of binding to
> the said hyperimmune serum-reactive antigen or of reducing
> or inhibiting the interaction activity of the said
> hyperimmune serum-reactive antigen or fragment thereof; the
> use of the said hyperimmune serum-reactive antigen or
> fragment thereof for the isolation and/or purification
> and/or identification of an interaction partner; a process
> for in vitro diagnosing a disease or a bacterial infection
> based on determining the presence of the said nucleic acid
> sequence or the presence of the said hyperimmune
> serum-reactive antigen or fragment thereof; the use of said
> hyperimmune serum-reactive antigen or fragment thereof for
> the generation of a peptide binding thereto i.e. an
> anticaline, for the manufacture of a functional nucleic acid
> i.e. an aptamer or a spiegelmer, or of a functional
> ribonucleic acid i.e. a ribozyme, an antisense nucleic acid
> or a siRNA.
>
> ———

Inventions 2-45: claims 1,2,5-11,14-37 (all partially)

> Idem as invention 1, but each of the inventions 2-45
> referring to one of the further SEQ ID Nos. mentioned in
> claim 1 together with its respective corresponing SEQ ID
> No. according to claim 11.
>
> ———

Inventions 46-121: claims 3-10,12,14-37 (all partially)

FURTHER INFORMATION CONTINUED FROM   PCT/ISA/ 210

Idem as invention 1, but each of the inventions 46-121
referring to one of the SEQ ID Nos. mentioned in claim 3
together with its respective corresponing SEQ ID No.
according to claim 12.

Inventions 122-133: claims 4-10,13-37 (all partially)

Idem as invention 1, but each of the inventions 122-133
referring to one of the SEQ ID Nos. mentioned in claim 4
together with its respective corresponing SEQ ID No.
according to claim 13.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 02083855	A	24-10-2002	CA	2444133 A1	24-10-2002
			WO	02083855 A2	24-10-2002
WO 02077021	A	03-10-2002	CA	2439431 A1	03-10-2002
			EP	1373513 A2	02-01-2004
			WO	02077021 A2	03-10-2002
WO 02059148	A	01-08-2002	AT	410798 B	25-07-2003
			AT	1302001 A	15-12-2002
			BR	0207067 A	15-06-2004
			CA	2436057 A1	01-08-2002
			CZ	20032201 A3	17-03-2004
			WO	02059148 A2	01-08-2002
			EP	1355930 A2	29-10-2003
			NO	20033364 A	24-09-2003